தமிழ்த் தரவுத்தொகுதியை தொடர்ப்பகுப்பாய்வு செய்தல்

(Chunking of Tamil Corpus)

சங்கரவேலாயுதன் இராசேந்திரன்
அமிர்தா விஷ்வவித்யபீடம், கோயம்புத்தூர்
rajushush@gmail.com

1 முன்னுரை

தொடர்ப்பகுப்பாய்வு (Chunking) என்பது வாக்கியங்களைத் தொடர்களாகப் (chunks) பகுத்தாய்தலின் முக்கியக் கூறாகும். பகுப்பாய்வு என்பது ஒரு இலக்கண அடிப்படையில் பனுவல்களின் தானியக்க ஆய்வுகளுடன் தொடர்புடையது. தொழில் நுட்ப அடிப்படையில் இது ஒரு பனுவலுக்கு தொடரியல் அமைப்பைத் தரும் செற்பாங்கைக் குறிப்பிடுகின்றது. இது பெரும்பாலும் அடிப்படை உருபன்-தொடரியல் வகைப்பாடுகள் கண்டுப்பிடிக்கப்பட்ட பின் நடைமுறைப்படுத்தப்படும். வேறுபட்ட இலக்கணங்கள் (சார்பு இலக்கணம், சூழல் இலக்கணம் போன்றவை) அடிப்படையில் பகுத்துக் குறித்தல் ஒன்றுடன் ஒன்று உயர்ந்த நிலைத் தொடரியல் உறவுகளில் கொண்டு வருகிறது. வாக்கிய நிலைப் பகுத்துக் குறித்தல் சொல்நிலைப் பகுப்பாய்விலிருந்து பெறப்பட்ட தகவலைப் பயன்படுத்த தானியக்கச் சூழல் அடிப்படையிலான மற்றும் சூழல் சுதந்திரமான தொடரியல் ஆய்வை உள்ளடக்குகின்றது. ஒரு பகுத்து அடையாளப்படுத்தப்பட்ட தரவுத்தொகுதி பகுப்பாய்வில் கிளைப்படங்களைப் பயன்படுத்துவதால் இது கிளைவங்கி (Tree bank) என்றழைக்கப்படுகின்றது. கிளை அமைப்பின் காட்சிப்படம் மிக அரிதாகவே தரவுத்தொகுதி அடையாளப்படுத்தலில் காணப்படும். பொதுவாக ஒத்த தகவல் அடையாளப்படுத்தப்பட்ட அடைப்புக்குறிகளின் குழுமத்தைப் பயன்படுத்தி உருப்படுத்தம் செய்யப்படும்.

எ.கா.

'Pearl sat on a chain' என்பது கிளை வங்கியில் பின்வருமாறு தோன்றும்.

[S [NP Pearl – NP/NP] [VP – Sat VVI] [PP on II]

[NP a – AJI] [Chair – NNI]

[NP] [PP] [VP] S]

இதில் உருபனியல், தொடரியல் தகவல் கீழ்க்கோடுகளால் சொற்களுடன் இணைக்கப்பட்டிருக்கும் உறுப்புகள் தொடர்களின் தொடக்கத்திலும் இறுதியிலும் திறக்கும் மற்றும் அடைக்கும் செவ்வக அடைப்புக் குறிகளால் அடையாளப்படுத்தப்படும்.

எல்லாப் பகுப்பாய்வு ஒழுங்கு முறைகளும் ஒன்றல்ல. முக்கியமான வேறுபாடுகள் பின்வருமாறு:

1. ஒரு ஒழுங்குமுறை பயன்படுத்தும் உறுப்பு வகைகளின் எண்ணிக்கை.
2. இந்த உறுப்பு வகைகள் ஒன்றுடன் ஒன்று இணைவதற்கு அனுமதிக்கப்படும் வகைகள்

இந்த வேறுபாடுகள் இருப்பினும் பெரும்பாலான பகுப்பாய்வுத் திட்டங்கள் சூழல் கட்டுப்பாடு இல்லாத தொடரமைப்பு இலக்கணத்தின் அடிப்படையால் அமைந்தது.

இந்த அமைப்பொழுங்குக்குள் முழு பகுப்பாய்வுத் (full parsing) திட்டம் வாக்கிய அமைப்பின் விரிவான ஆய்வைத் தருவதை நோக்கமாகக் கொண்டது. குறுகிய பகுப்பாய்வுத் திட்டம் (Skeleton parsing scheme) தொடரியல் உறுப்பு வகைகளில் குறைந்த நிலையிலான வேறுபடுத்தலைப் பயன்படுத்துகின்றது. மேலும் சில உறுப்பு வகைகளின் அக அமைப்பை விட்டுவிடுகின்றது. பகுப்பாய்வு பெரும்பாலும் மனித ஆய்வில் பின் திருத்தம் செய்யப்படுகின்றது. ஏனென்றால் தானியங்குப் பகுப்பாய்வு சொல்வகை அடையாளப்படுத்தலை விட வெற்றி குறைந்தது. முழு மனித இயக்கப்

பகுப்பாய்வின் குறைப்பாடு பகுப்பாய்வு, தரவுத்தொகுதி பிரித்தல் இவற்றில் ஈடுபடும் ஆய்வாளரின் ஒழுங்கின்மையாகும். இதை ஈடுக்கட்ட விரிவான வழிமுறைகள் தரப்படுகின்றன. இருப்பினும் பல்பொருள் கோள் சாத்தியமாகும்போது மயக்கம் ஏற்படலாம். கிளை வங்கிகள் இயற்கை மொழிகளுக்கு அதன் அமைப்பின் வேறுபட்ட நிலைகளில் (சொல்நிலை, வாக்கியநிலை, தொடர்நிலை, செயல்பாடு–பங்கெடுப்பாளர் நிலை) அடையாளப்படுத்தும் நிலை தரும் மொழி மூலவளமாக அமைகின்றது. கிளை வங்கிகள் இயற்கைமொழி ஆய்வு தரவு உந்தல் அடிப்படைகள், மனித மொழி தொழில் நுட்பங்கள், இலக்கணப் பிரித்தெடுத்தல், பொதுவாக மொழியியல் ஆய்வுகள் இவற்றின் உருவாக்கத்திற்கு முக்கியமானதாகும்.

அடிப்படையில் தொடர்ப்பகுப்பான்/தொடர்ப்பகுப்பாய்வி (chunker) வாக்கியத்தை மேலுறல் செய்யாத முக்கியமான தொடர்களாகப் பிரித்து அதற்குப் புலக்குறிப்பு (label) தருகிறது. தொடர்ப்பகுப்பான் அவற்றின் துல்லியமான வெளிப்யீடு அடிப்படையிலும் எவ்வாறு தொடர்ப்பகுதிகள் விளக்கப்படுகிறது என்பதன் அடிப்படையிலும் வேறுபடுகின்றன. பல தொடர்ப்பகுப்பான்கள் எளிய தொடர்களைப் பகுத்தலைவிடக் கூடுதலாகச் செய்கிறது. பிற பெயர்த்தொடர்களை மட்டும் கண்டுபிடிக்கிறது. தொடர்ப் பகுப்பான்கள் சொல்வகை அடையாளப்படுத்தலுக்கும் (parts-of-speech tagging), முழுப் பகுப்பாய்வுக்கும் (full parsing) இடையில் வருகிறது. தனிப்பட்ட தொடர்ப் பகுதிகளின் (chunks) அமைப்பை விளக்குவது எளிது. ஆனால் அவற்றிற்கிடையிலான உறவுகளை விளக்குவது கடினம். இது தனிப்பட்ட சொல் வகுப்புகளைச் சார்ந்திருக்கிறது. இவ்வாறு தொடர்களைப் பகுத்ததில் தற்பொழுது இருக்கின்ற முறைக்கும் முன்மாதிரி வெளிப்யீட்டு ஆய்வுக்கும் இடையில் உள்ள சலுகையான முறையாகும். தொடர்ப் பகுப்பான்கள் வாக்கியங்களைப் பிரித்து அடையாளப்படுத்துகிறது. பெரும்பாலான தொடர்ப் பகுப்பான்கள் சொல்வகை அடையாளப்படுத்தலில் (parts-of-speech (POS)) உள்ள தகவலைப் பயன்படுத்துகிறது. பிற உண்மையான சொற்களைப் பார்க்கிறது.

சொல் வகைப்பாட்டு அடையாளப்படுத்தி (parts-of-speech (POS) tagger)சொல் மற்றும் சூழலுக்கான தகவலைப் பயன்படுத்தி இம்மயக்கங்களை நீக்குகின்றது. சொல் வகைப்பாட்டுத் தகவல் பகுத்துக்குறித்த அமைப்பில் உறுப்பின் நிரல் அல்லது இடத்தைக் கண்டுகொள்ள தேவைப்படுகிறது. பண்புக்கூறு தகவல்கள் நிரப்பியையும் உடன்பாட்டையும் பரிசோதிக்கப் பயன்படுகிறது.

தமிழில் ஒத்தறி அடிப்படையில் சொல்நிரல் சுதந்திரமானது. தொடரமைப்பு விதிகளின் எல்லா விதமான வரிசை மாற்றங்களுக்கும் வேண்டி எழுதப்பட வேண்டும் என்று அவசியப்படும் இவ்வகங்களின் காரணமாக இப்பகுப்பான்கள் தேவையில்லாமல் விதிகளின் எண்ணிக்கையைப் பெருக்கிக் கொள்ளும். பகுப்பான் மீது கூடதலான சுமை புகுத்தப்படும். தற்போதைய ஆய்வில் பகுப்பாய்வுக்கு முன் வாக்கியத்தை ஒரு குறிப்பிட்ட நிலை பேறாக்கம் செய்யும் நடத்தை பின் பற்றப்படுகிறது.

வினைதான் வாக்கியத்தின் முக்கிய உறுப்பாகும். வினைகள் அவற்றின் பொருண்மை இயல்பு அடிப்படையில் வேறுபட்ட பங்களிப்பாளர் அமைப்பை ஏற்கும். இப்பங்கெடுப்பாளர் அமைப்புகள் பெயர்த்தொடர்களின் வேற்றுமை உருபுகளை விளக்கும். அவை வினைக்கு நேரடியான மற்றும் நேரடியல்லாத செயப்படுபொருளாக அமையும். பகுப்பான் வாக்கியத்தில் வினைக்குப் பொருத்தமான பங்கெடுப்பாளர் அமைப்பை பரிசோதிப்பதைக் கவனிக்க வேண்டும். மேலும், தமிழில் எழுவாயின் பால்-எண்-இட பண்புக்கூறுகள் பயனிலையுடன் இயைபுடையதாக இருக்க வேண்டும். ஒரு பகுத்துக்குறிப்பான் வாக்கியத்தைப் பகுத்துக்குறிக்கும்போது வாக்கியத்தைச் சரியாக மதிப்பிட இவ் இயைபு பண்புக்கூறைப் பரிசோதிக்க வேண்டும்.

2. பகுப்பாய்வு (parsing)

பகுப்பாய்வு இயந்திரமொழிபெயர்ப்பு மற்றும் தகவல் பிரித்தெடுத்தல் போன்ற பயன்பாடுகளில் முக்கியமான செயல்பாடாகும். சில பயன்பாடுகள் வாக்கியங்களின்

முழுமையற்ற பகுத்தாய்தலை (shallow parsing) வேண்டுகிறது. ஆங்கிலம் மற்றும் பிற ஐரொப்பிய மொழி வாக்கியங்களின் பகுப்பாய்வுகுப் பல முயற்சிகள் மேற்கொள்ளப்பட்டுள்ளன. இந்திய மொழிகளில் இந்தி மற்றும் வங்காள மொழிகளுக்கு பகுப்பான்கள் இருந்தாலும் தமிழைப் பொறுத்தவரையில் கூறத்தக்க பகுப்பான் இல்லை. தமிழில் பல கோட்பாட்டு மொழியியல் அடிப்படையிலான தொடரியல் ஆய்வுகள் செய்யப்பட்டுள்ளன. இருப்பினும், தமிழ் வாக்கியங்களின் பகுப்பாய்வுக்கு வேண்டிய கணினியியல் சார் சட்டகத்தை உருவாக்க அச்சிட்டு வெளியிடப்பட்ட புத்தகங்கள் இல்லை. தற்போதைய ஆய்வு தமிழ் வாக்கியங்களின் பகுப்பாய்வுக்குக் கணினி அடிப்படையிலான இலக்கணச் சட்டகம் உருவாக்குவதில் கவனம் செலுத்தப்படுகின்றது.

3 ஆழமிலாப் பகுப்பாய்வு (shallow parsing)

நாம் தொடரியல் என்ற கலைச்சொல்லை மரபுப் பகுப்பாய்விலிருந்து கிடைக்கும் வெளிப்யீட்டை விடக் குறைவான நிறைவுடைய ஆய்வைக் குறிக்கும் பொதுவான சொல்லாகப் பயன்படுத்துகிறோம். ஆய்விலிருந்து கிடைக்கும் வெளிப்யீடு தொடரமைப்புக் கிளைப்படம் அல்ல. ஒரு பகுப்பான் பெயர்த்தொடர்கள் போன்ற பகுக்கப்பட்ட உறுப்புகளை வாக்கியத்தில் அவற்றின் அக அமைப்பையும் செயல்பாட்டையும் குறிப்பிடாமல் அடையாளம் காணும். மற்றொரு வகையான பகுப்பான் முதன்மை வினை மற்றும் அவற்றின் நேரடியான பங்கெடுப்பாளர் போன்ற சொற்களின் சில செயல்பாட்டுப் பங்களிப்பை அடையாளம் காணும். பகுப்பாய்வு ஒழுங்குமுறை உருபயனியல் ஆய்வின் மீதும் பொருண்மை மயக்க நீக்கலின் மீதும் இயல்பாக வேலை செய்யும். இதன் அடிப்படை நோக்கம் தலைச்சொல் (head word) உருபனியல் தகவல் மற்றும் சொல்வரிசை அமைப்பு இவற்றில் இருந்து சாத்தியமான தொடரியல் அமைப்பைப் (syntantic structure) பெறுவதாகும். முன்மாதிரியாக பகுப்பாய்வுத் தொடர்களையும் தலை-அடை (head-medifier) உறவுகளையும் கண்டுபிடிப்பதை நோக்கமாகக் கொள்கிறது. பெரிய உரை தரவுத்தொகுதிகளின் (corpus) பயன்பாட்டில் ஆழமில்லாப் பகுப்பான் பயன்படுத்தப்படுகிறது. பெரும்பாலும்

பகுப்பாய்விற்கு எல்லாச் சிக்கல்களுக்கும் தீர்வு காண்பதற்கு போதிய திறன் இல்லாவிட்டால் பகுதியான ஆய்வுகள் அனுமதிக்கப்படும். சர்ச் (Church.1988) பொருத்தமான அடைப்புக் குறிகளைச் செருகி எளிய பெயர்த் தொடர்களை அடையாளம் காணும் புள்ளியல் (Stochastic) வழியமைப்பை வடிவமைத்துள்ளார். ஒப்னெய் (Obney, 1991) உளமொழியியல் சான்றுகள் அடிப்படையிலும், நடைமுறை பயன்பாடுகள் அடிப்படையிலும் ஆழமில்லாப் பகுப்பாய்வின் முக்கியத்துவத்தை முதலில் விவாதித்தவர் என்று மதிக்கத்தக்கவர் ஆவார். அவர் தமது அணுகுமுறையில் ஆழமில்லாப் பகுப்பாய்வைப் பெறுவதற்காக முற்றுநிலை மாறிகளைப் பயன்படுத்தினார். ஆழமில்லாப் பகுப்பான் சொல்வகைப்பாடு அடையாளப்படுத்தல், தொடர்களை வரையறுத்தல், தொடர்களைக் கண்டுபிடித்தல் என்ற முன்மாதிரி தொகுப்புகளை உள்ளடக்கியது.

3.1 சொல்வகை அடையாளப்படுத்தல் (Parts of Speech tagging)

ஒரு சொல்லும் அதன் சூழலும் தரப்படுகையில் சொல்வகை அடையாளப்படுத்தி அச்சொல்லின் சரியான உருபனியல்-தொடரியல் வகுப்பை (பெயர், வினை கொண்ட பகுப்பை) உறுதிசெய்யும் சொல்வகை அடையாளப்படுத்தி சொல்வகை அடையாளப்படுத்தல் (Parts of Speech tagging) இயற்கைமொழி ஆய்வில் மிக அறியப்பட்ட சிக்கலாகும். இதில் இயந்திர கற்றல் அணுகுமுறைகள் (machine language approaches) பொதுவாகப் பயன்படுத்தப்படுகிறது.

3.2 தொடர்ப் பகுப்பான் (chunker)

சொற்களும் அவற்றின் உருபனியல்-தொடரியல் வகுப்புகளும் தரப்படுகையில் எந்தச் சொற்கள் தொடர் உறுப்புகளாகக் (பெயர்த்தொடர், வினைத்தொடர், நிரப்பியத்தொடர் போன்றவைகளாக) குழுமப்பட வேண்டும் என்று தொடர்பகுத்தல் (Chunking) செயல்பாட்டால் தீர்மானிக்கப்படுகிறது.

தொடரின் வரையறை விளக்கத்துடன் பின்வரும் கேள்வி தொடர்புள்ளது. தொடர் எதைக் கொண்டிருக்கும்? ஒரு மாதிரி தொடர் ஒரு தனி நிலைப் பொருளடக்கச் சொல்லைச் சுற்றி ஒரு கூட்டம் செயல் பாட்டுச் சொற்களைக் கொண்டிருக்கும் (Abney, 1991). தொடர்கள் இயல்பாக ஒன்றுக்கொன்று இயைபுள்ள சொற்களின் குழுமமாக எடுத்துக்கொள்ளப்படுகின்றது.

அடுத்த கேள்வி 'நம்முடைய தேவைக்காக இந்த இயைபுள்ள சொற்களின் குழுவின் எல்லைகளை எவ்வாறு வரையறை விளக்கம் செய்யவியலும்?' என்பதாகும். எடுத்துக்காட்டாக, பின்வரும் இணைகளில் எது சரியான தொடர் உறுப்பு? என்ற ஐயம்.

1. ((மேசையின் மேல்)) அல்லது
2. ((மேசையின்)) ((மேல்))

3.3 தொடர் உறவைக் கண்டுபிடித்தல்

ஒரு வாக்கியத்தில் உள்ள தொடர் உறுப்புகள் தரப்படுகையில் முக்கிய வினையுடன் அவற்றிற்குள்ள உறவுகள் (எழுவாய், பயனிலை, இடம் போன்றவை) தீர்மானிக்கப்பட வேண்டும் என்ற நிலையில் உறவைக் கண்டுபிடித்தல் (Relation finding) முக்கியமான செயல்பாடாகும்.

ஆழமில்லாப் பகுப்பாய்விகள் இயற்கை மொழி முழுவதையும் கையாள வேண்டி இருப்பதால் அவை பெரிதாகவும் பெரும்பாலும் ஆயிரக்கணக்கான விதிகளையும் கொண்டிருக்கின்றன. இவ்விதிகளில் பல விதிவிலக்கானவை என்பதன் அடிப்படையில் இவை வலிமை குறைந்தவையாக இருக்கின்றன. ஆழமில்லாப் பகுப்பாய்விகள் இயந்திர கற்றல் அடிப்படையிலான தொழில்நுட்பங்களைப் (techniques) பயன்படுத்தித் தானியக்கமாக உருவாக்கப் படுகின்றன.

4 பெயர்த்தொடர் பகுப்பு (NP chunking)

பெயர்த்தொடர்களைப் பகுத்தல் வாக்கியங்களிலிருந்து பெயர்த் தொடர்களைப் பிரித்தெடுப்பதைக் குறிப்பிடுகின்றது. பகுத்தாய்வதைவிட பெயர்த்தொடர் பகுத்தல் எளிதாக இருந்தாலும் ஒரு துல்லியமான மற்றும் திறமைவாய்ந்த பெயர்த்தொடர் பகுத்தல் பல பயன்பாடுகளில் பயன்படுத்தப்படுகின்றது என்ற உண்மையால் மிக முக்கியத்துவம் வாய்ந்ததாகக் கருதப்படுகின்றது.

பெயர்த்தொடர் பகுப்பைப் பனுவல்களைப் பகுத்தாய்வதற்கு முன் முன்பகுப்பாய்வுக் கருவியாகப் (pre-processing tool) பயன்படுத்த இயலும். இயற்கை மொழியின் கூடுதலான பொருண்மை மயக்கம் காரணமாக பனுவலின் சரியான பகுப்பாய்வு மிகக் கலவைத் தன்மையானதாக (சிக்கலானதாக) இருக்கின்றது. இந்நேர்வுகளில் தொடர்களைப் பகுத்தலை இம்மயக்கங்களை ஓரளவு தீர்ப்பதற்கான முன்பகுப்பாய்வுக கருவியாகப் பயன்படுத்த இயலும். இப்பயன்பாட்டில் தொடர் பகுத்தலைச் சொற்களுக்குப் பதிலாகத் தொடர் பகுதிகளைச் சார்ந்திருக்கும் ஆவணங்களிலிருந்து தரவை மீட்பதற்குப் பயன்படுத்த இயலும். குறிப்பாகப் பெயர்களும் பெயர்த்தொடர்களும் தகவல் மீட்புக்கும் பிரித்தெடுத்தலுக்கும் மிகவும் பயனுள்ளதாகும். இயந்திர மொழிபெயர்ப்பு மீதான சமீபகாலச் செயல்பாடுகள் பயனுள்ள மாற்ற அமைப்பொழுங்குகளை (transfer patterns) ஆக்குவதற்கு இரு மொழிகளிலுள்ள பனுவல்களை (இணை தரவுத்தொகுதிகளை) (parallel corpora) பயன்படுத்துகின்றது. பெயர்த்தொடர்கள் இணை தரவுத்தொகுதிகளில் உள்ள பனுவல்களைப் பொருத்தும் பயன்பாடுகளில் பயன்படுகின்றது. இணை தரவுத்தொகுதிகளில் உள்ள வாக்கியங்களைத் தொடர் பகுப்புத் தகவல்களைப் பயன்படுத்தியும் மூல மொழியிலுள்ள தொடர்ப்பகுதிகளை இலக்கு மொழியிலுள்ள தொடர் பகுதிகளுடன் உறவுபடுத்தியும் பொருத்த வரைவு செய்யலாம். மேலும் பகுக்கப்பட்ட பெயர்த் தொடர்களைத் தரவின் தரமான பகுப்பாய்வில் தேவைப்படாத பிற பயன்பாடுகளிலும் பயன்படுத்த இயலும்.

5. தமிழ்த் தரவுத் தொகுதியின் தொடர்ப்பகுப்பாய்வு

5.1 தமிழ்த் தொடரியல் அமைப்பு

தமிழைப் பொறுத்தவரையில் அது ஒரு ஒட்டுக்கள் நிறைந்த மொழியாகும். இவ்வொட்டுக்கள் வேர்ச்சொற்களில் சேர்க்கப்பட்டு வேறுபட்ட சொல் வடிவுகளை உருவாக்கும். ஏனவே தமழுக்கு ஒரு உருபனியல் ஆய்வியை உருவாக்குவது ஒரு சிக்கலான செயல்பாடாகும். பல வாக்கியங்களுக்கு உருபனியல் பகுப்பாய்வியைப் பயன்படுத்தி சொற்களின் அல்லது ஒரு சொல்லின் சொல் வகைப்பாட்டையும் (Parts of speech (POS)) பண்புக்கூறு தகவலையும் அடையாளம் காண இயலும். இருப்பினும், சில சொற்களுக்கு மயக்கமுள்ள சொல்வகைப்பாடும் இலக்கண பண்புக்கூறுகளும் இருக்கும்.

தமிழ்த் தொடரியல் பகுபாய்வுக்கு அடிப்படையாக அமையும் தமிழ்த் தொடரியல் குறித்து இங்கு விளக்கப்படும். தமிழ் வாக்கியங்களின் வகைகள் பற்றியும், அவற்றின் அமைப்பு குறித்தும் விளக்கப்படும். இப்பகுதி அகத்திலிங்கம் அவர்கள் (2002) எழுதி வெளியிட்டுள்ள "தமிழ்மொழி அமைப்பியியல்" என்ற நூலின் அடிப்படையில் அமைந்ததாகும்.

தமிழ்மொழி வேறுபட்ட சொல் வரிசை அமைப்பைக் கொண்ட தொடர்களைக் கொண்டு வாக்கியங்களை அமைக்கத்தக்க வகையில் அமைந்த மொழியாகும். சொற்களுக்கு இடையிலான தொடர்பினைப் பயன்படுத்தி வாக்கியங்களை இலக்கணரீதியாக அமைக்கிறோம். இலக்கணங்கள் வேறுபட்ட வார்த்தைகளுக்கு இடையே தொடர்பினை உருவாக்கிப் பொருள்தரும் வாக்கியங்களை உருவாக்க உதவுகிறது. இவ்வாறு சொற்களைக் கொண்டு தொடர்களையும் தொடர்களைக் கொண்டு வாக்கியங்களையும் உருவாக்கும் விதிகளைக் கூறுவது தொடரியலாகும்.

ஒன்றுக்கு மேற்பட்ட சொற்கள் அத்தொடர்கள் இணைந்து உருவாகும் வாக்கியங்கள், அவ்வாக்கியங்கள் இணைந்து உருவாகும் கலப்பு, கூட்டு வாக்கியங்கள் ஆகியவற்றின் கட்டமைப்பையும் அவை தரும் பொருண்மையையும் ஆராய்வது தொடரியல் எனப்படும். பொதுவாகத் தொடர்களை ஆராய்வது தொடரியல். அது தொடராக இருக்கலாம் அல்லது வாக்கியங்களாக இருக்கலாம். வாக்கியம் என்பதும் ஒருவகைத் தொடரே.

5.1.1 தொடர்கள்

தமிழ் வாக்கியங்களைப் தலைமையாய் வரும் சொல்வகைப்பாடுகளின் அடிப்படையில் பெயர்தொடர், வினைதொடர், பெயரடைத் தொடர், வினையடைத் தொடர், பின்னுருபுத்தொடர் என வகைப்படுத்தலாம்.

5.1.1.1 பெயர்த்தொடர்

பெயர்த்தொடர் என்பது ஒரு பெயர்ச்சொல்லைத் தலைமைப் பெயராகக் கொண்ட தொடர் ஆகும். தலைமைச்சொல் பெயராகவோ ஆக்கப்பெயராகவோ (எகா.மயக்கம்) தொகையாகவோ (எகா. பணப்பெட்டி) வினைப்பெயராகவோ, மாற்றுப்பெயராகவோ (எகா.அவன், நான்,நீ போன்று) தொழில் பெயராகவோ (வந்தது, வருவது, வருதல், வந்தமை, வந்தவன் போன்று) இருக்கலாம். பெயர்த்தொடரில் பெயருக்கு முன் அடைகொளி அடை, எண்ணடை, பெயரடை, அளவையடை போன்ற அடைகளும் பெயரெச்சத் தொடரும் வரலாம்.

எ.கா.

பையன், அந்தப் பையன், பெரிய பையன், இரண்டாவது பையன், இரண்டு பையன்கள், என்னுடைய பையன், நேற்று வந்த பையன்.

5.1.1.2 வினைத்தொடர்

ஒரு வினைச்சொல்லைத் தலைமைச் சொல்லாகக் கொண்டு உருவாகும் தொடர் வினைத்தொடர் எனப்படும். வினைத்தொடரில் வினைக்கு முன்னர் வேற்றுமைத்தொடர், பின்னுருபுத்தொடர், வினையடை, வினையெச்சத்தொடர் என்பன வரலாம்.

எ.கா.

விழுந்தேன், செய்தேன், அங்கு வந்தேன்
நன்றாக விளையாடினேன், வந்து பார்த்தேன்

போன்றவை வினைத்தொடர்களாகும்.

5.1.2 வாக்கிய வகைகள்

வாக்கியங்களைப் பல்வேறு அடிப்படையில் பல வகைகளாகப் பிரிப்பர் இலக்கணயலார். அமைப்பினை அடிப்படையாக வைத்துத் தமிழ் வாக்கியங்களை வினைப்பயனிலை வாக்கியங்கள் என்றும், பெயர்ப் பயனிலை வாக்கியங்கள் என்றும் பிரிப்பர். இதுபோன்று வாக்கியங்களின் சேர்க்கையை அடிப்படையாகக் வைத்து தனி வாக்கியம் அல்லது எளிய வாக்கியம் (Simple sentence) என்றும், கலவை வாக்கியம் (Complex sentence) என்றும் என்றும் பிரிப்பர். வெவ்வேறு வினைகளின் பண்பையும் அதனால் அடையும் மாற்றங்களையும் வைத்து செய்வினை வாக்கியம் (active voice) என்றும், செயப்பாட்டு வினை வாக்கியம் (passive sentence)என்றும் பிரிப்பது உண்டு. இவை மட்டுமின்றி வேறொரு பண்பினை ஒட்டி இலக்கணங்கள் வாக்கியங்களை வினா வாக்கியம் (interrogative sentence) என்றும், செய்தி வாக்கியம் (affirmative sentence) என்றும் பிரிப்பர். இவை மட்டுமல்லாமல் இலக்கணவியலார் ஏவல் வாக்கியம், செய்தி வாக்கியம் என்றும் பிரிப்பர். தமிழில் காணப்படும் வாக்கியங்களை 1. வினைப் பயனிலை

வாக்கியம், (கண்ணன் வந்தான்), 2. பெயர்ப் பயனிலை வாக்கியம் (இராமன் ஒரு ஆசிரியர்), 3. பெயரடை பயனிலை வாக்கியம் (இராமன் நல்லவன்), 4. உடைமைப் பயனிலை வாக்கியம் (இராமன் என்னுடையவன், புத்தகம் என்னுடையது) என நான்காகப் பிரித்து இனம் காண்பர் சிலர் (கோதண்டராமன்,1997) .

தொடரியலில் சில முக்கியமான அமைப்புகளை மட்டுமே குறித்துச் சொல்லுகின்ற நிலையில் இதன் கண் ஒருசில அமைப்புகள் மட்டுமே எடுத்துக் கூறப்படுகின்றன.

5.1.2..1 அடிப்படை வாக்கிய வகைகள்

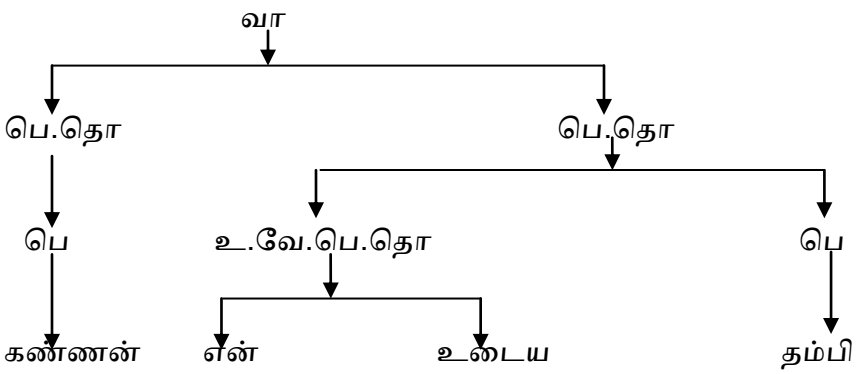
தமிழில் காணப்படும் அடிப்படையான வாக்கியங்கள் 1. பெயர்ப் பயனிலை வாக்கியம் என்றும், 2. வினைப் பயனிலை வாக்கியம் என்றும் இரு வகைகளாகப் பிரிக்கலாம். இவ்விரண்டு வாக்கிய வகைகளிலும் பெயர்த்தொடர் எழுவாய்களாக வரும் இந்நிலையில் இவற்றை

1.பெயர்த்தொடர் + பெயர்த்தொடர் (NP + NP)

2.பெயர்த்தொடர் + வினைத்தொடர் (NP + VP)

என்ற அமைப்பைக் கொண்டுள்ளவைகளாகப் பிரிக்கலாம்.

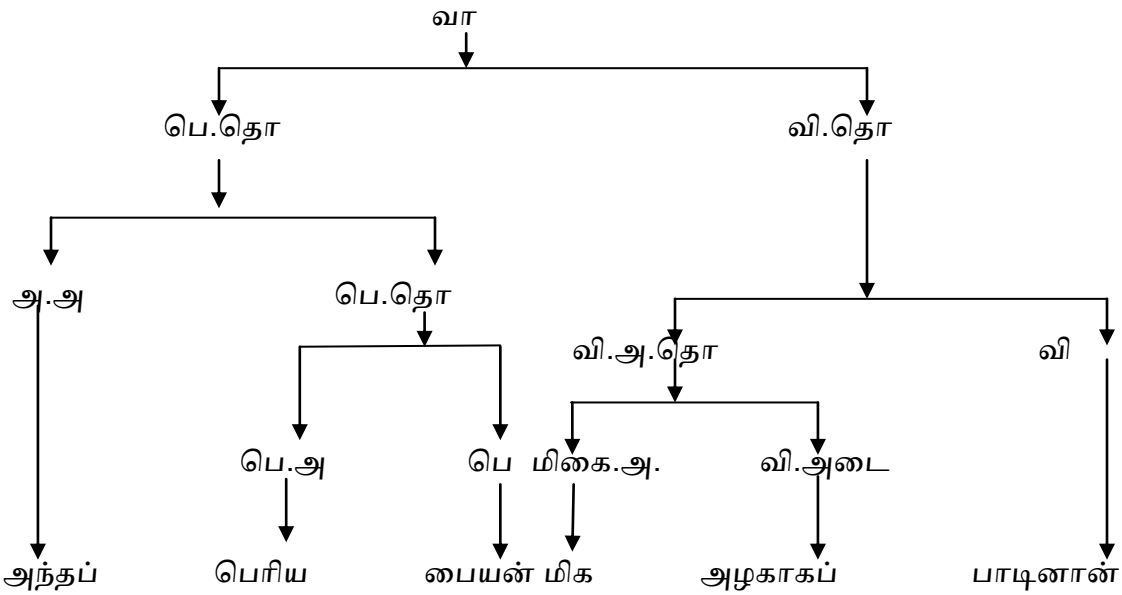
ஒரு பெயர்ப் பயனிலைத் தொடர் வாக்கியம் எவ்வாறு உள்ளது என்பது கீழே உள்ள கிளைப்படத்தில் காட்டப்பட்டுள்ளது.



சுருக்கவிரிவு: வா= வாக்கியம்பெ.தொ. = பெயர்த்தொடர், பெ = பெயர், உ.வே.பெதொ = உடைமை வேற்றுமைப் பெயர்த்தொடர்.

மேலே 'கண்ணன் என்னுடைய தம்பி'என்ற பெயர்ப் பயனிலை வாக்கியத்தின் அமைப்பு காட்டப்பட்டுள்ளது.

பின்வரும் வினைப் பயனிலை வாக்கியம் ஒரு வினையையும் வினையடைத் தொடரையும், மிக என்ற மிகையடையும் கொண்டு வந்துள்ளது.



அ.அ. =அடைகொளி அடை, வி.அ.தொ.=வினையடைத்தொடர், வி.அ. = வினையடை, மிகை.அ. =மிகையடை

5.1.2..2 வினா வாக்கியங்கள்

வினா வாக்கியம் என்பது தெரியாத ஒன்றை மற்றொருவரிடம் கேட்டுத் தெரிந்து கொள்ளுதற்காகப் பயன்படுத்தப்படும் வாக்கியம் ஆகும். வினாவில் அடிவினா என்றும், ஒட்டுவினா என்றும் இருவகை உள்ளன. அடிவினா யார், என்ன போன்ற வினாச் சொற்களையும் ஒட்டுவினா ஆ என்ற ஒட்டையும் (எ.கா. அவனா? இவனா?) கொண்டிருக்கும்.

வினா வாக்கியங்களும் அவற்றின் விடையாக வரும் செய்தி வாக்கியங்களும் ஏகதேசம் ஒரே அமைப்பைக் கொண்டு விளங்கும்.

ஆம் இராமன் வந்தான்.
அவன் நேற்று வந்தான்.

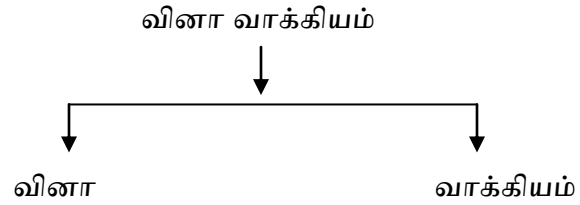
என்ற செய்தி வாக்கியங்களும்,

இராமன் வந்தான். இராமன் வந்தானா?
அவன் நேற்றா வந்தான்?

என்ற ஒட்டு வாக்கியங்களும் ஏகதேசம் ஒரே அமைப்பைக் கொண்டவை. ஒன்றில் வினா ஒட்டு உள்ளது. செய்தி வாக்கியத்தில் அது இல்லை. இதே நிலைதான் அடிவினா வரும் வாக்கியங்களிலும் காணப்படுகின்றது.

இராமன் வந்தான். இராமன் என்று வந்தான்?
இராமன் ஒரு ஆசிரியர். இராமன் யார்?
அவன் நன்கு செய்தான் அவன் எப்படிச் செய்தான்?

போன்ற இணை வாக்கியங்களும் ஒரே மாதிரியான அமைப்பையே கொண்டுள்ளன. ஒன்றில் செய்தி (நேற்று) மற்றொன்றில் வினா (என்று) இதனால் மொழியியலாளர்மள் வினா வாக்கியங்களையும் அதை ஒட்டி வரும் செய்தி வாக்கியங்களையும் ஒரே அமைப்பின் கீழ்க் கொண்டு வந்து வினா வாக்கியத்தைப் பிரித்துக்காட்டும் நிலையில் 'வினா' என்ற முன்னிலையை இணைத்து வேறுபடுத்துவர். இதனைக் கீழ வருமாறு காட்டலாம்.



ஒட்டு வினாவும், அடிவினாவும் ஒன்று வந்த இடத்தில் மற்றொன்று வருவதில்லை. இதனைக் கீழ்வருமாறு காட்டலாம்.

வினா-ஒட்டுவினா, அடிவினா

ஒட்டுவினாப் பொருள்

ஒட்டுவினா ஆம் அல்லது இல்லை என்ற பதில்களை எதிர்பார்ப்பது. முதலில் உள்ளது கேட்போரின் உள்ளத்தில் உள்ளதை உறுதி செய்வது . இரண்டாவது உள்ளதை மறுப்பது.

இராமன் வந்தான்.
இராமன் வந்தானா?

எனத் தமிழில் வருவது காணத்தக்கது. முதல் வினாவின் விடையாக ஆம் (இராமன் வந்தான்) என்றோ இல்லை (இராமன் வரவில்லை) என உறுதி செய்தோ அல்லது

மறுத்தோ விடை அமையலாம். இதுபோன்றே அடுத்த வினாவிற்கும் பதில்கள் அமையலாம்.

தமிழில் ஒட்டுவினா, வினா-ஆ என்ற உருபனால் காட்டப்படுகின்றது. பொதுவாக இது வாக்கியங்களில் ஒரே ஒரு இடத்தில் வரும் பல்வேறு இலக்கணக் கூறுகளில் ஏதாவது ஒன்றின் பின்னால் வந்து ஒட்டு வினாப்பொருளைக் காட்டும். இவ்வாறு ஏதாவது ஒன்றின் பின்னால் வருவதை வினாவின் வீச்சு (ளஉடிப்ந) என்பர் மொழியியலாளர்கள்.

இராமன் + வினா + வந்தான் >இராமன் வந்தான்.

இராமன் வந்தான் + ஆ >இராமன் வந்தானா?

என்று வருவது குறிப்பிடத்தக்கது.

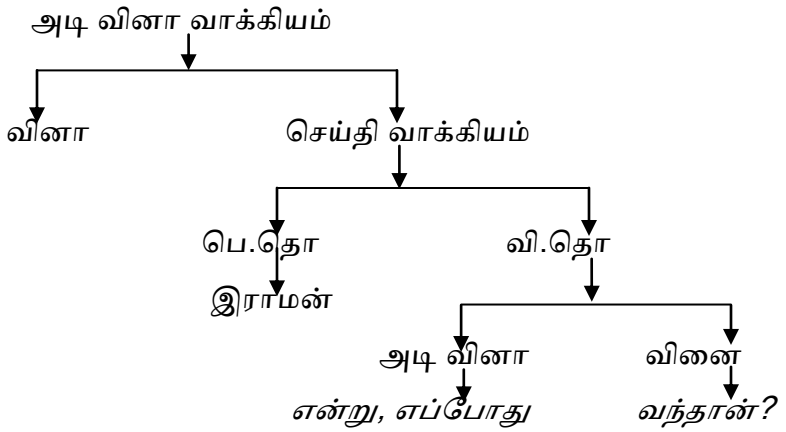
அடி வினா

அடி வினா வாக்கியங்கள் 1.யார், எவர், எவன், என்ன, எது, எவை போன்ற வினாப் பெயர்களையும், 2.எந்த, எத்தனையாவது, எத்தனை, எவ்வளவு, எப்பேர்பட்ட அல்லது எப்படி உள்ள (எத்தகைய, எம்மாதிரி போன்ற) என்ன போன்ற வினாப் பெயரடைகளையும், 3.எத்தனையாவது, எங்கு, எப்படி, எப்போது போன்ற வினா வினையடைகளையும் உள்ளடக்கியவை. இவற்றுள் சில ஒன்றுக்கு மேற்பட்ட நிலைகளிலும் (என்ன, எப்பொழுது) வரும்.

இராமன் நேற்று வந்தான்.

இராமன் என்று வந்தான்?

போன்ற வாக்கியங்கள் ஒரே அமைப்பைக் கொண்டுள்ள நிலையில் 'நேற்று' 'என்று' போன்றவை ஒரே அமைப்பு கொண்ட வாக்கியங்களில் ஒரே இடத்தில் வருபவை ஆகும். எனவே அடி வினா வாக்கியங்களின் அமைப்பைக் கீழ்வருமாறு காட்டலாம்.



கண்ணன் என்று/எப்போது வந்தான்

என்ற ஒரே அமைப்பைக் கொண்டுள்ள நிலையில் அடி வினா என்ற புனை நிலையில் (hypothetical) கொள்ளப்பட்டது கெட அடி வினா வாக்கியம் உருவாக்கப்பெறுகிறது.

5.1.2.3 செய்வினை, செயப்பாட்டு வினை வாக்கியங்கள்

வினாப் பாகுப்பாடு பற்றிக் கூறும்போது வினைகள் செய்வினை என்றும் செயப்பாட்டு வினை என்றும் பகுக்கப்படும் என்று கூறப்பட்டது. செய்வினைப்யன் செயவென் வாய்பாட்டு வினஇயெச்சத்தின் பின்னர் படு அல்லது பெறு என்ற துணை வினைகளை இணைக்கும்போது உண்டாகும் வினைதான் செயப்பாட்டு வினை எனவும் கூறப்படுகிறது. அவ்வாறு உள்ள செயப்பாட்டு வினையைப் பயனிலையாகக் கொண்ட வாக்கியங்கள் செயப்பாட்டுவினை வாக்கியங்கள் ஆகும்.

செயப்படுபொருளை எழுவாயாக ஆக்கியும் செய்பவனை அல்லது மருத்தாவை ஆல் வேற்றுமையாக ஆக்கியும் கூறப்படுவது செயப்பாட்டு வினை வாக்கியம் (Passive Sentence).

இராஜா இராமனை அடித்தான்

என்பது செய்வினை வாக்கியம். இங்கு செய்பவன் எழுவாயாகவும், செய்யப்படுபவன் ஐ வேற்றுமையாகவும் உள்ளது. இதனை உரிய இலக்கண மாற்றத்தால் மாற்றி

இராமன் இராஜாவால் அடிக்கப்பட்டான்

என்ற நிலையில் உருவாக்கம் பெறுகின்றது. இங்குச் செய்வினை வாக்கியம் அடிப்படை வாக்கியமாகவும் (புகைநிலை), செயப்பாட்டு வினை வாக்கியம் வருவிக்கப்பட்ட வாக்கியமாகவும் (புறநிலை) காணப்படுகின்றன. இவற்றிற்கு இடையே காணப்படும் மாற்றத்ததைக் கீழ்வருமாறு காட்டலாம்.

பெ.தொ.1+பெ.தொ.2_ஐ+வினை+காலம்+விகுதிகள்

செ.பா

பெ.தொ.2+பெ.தொ.1-ஆல்

வினை. செய.எ.எச்சம்+பகு+காலம்+விகுதிகள்

என்ற நிலையில் உருவாக்கம் பெறுகின்றது.

5.1.2.4 கலப்பு வாக்கியம்

ஒரு வாக்கியத்தின் உள்ளே இன்னொரு வாக்கியத்தை முன்னதன் பகுதியாக அல்லது உறுப்பு வாக்கியமாக இணைக்கும்போது உருவாகும் வாக்கியம் கலப்பு வாக்கியம் எனப்படும். எந்த வாக்கியம் தலைமை வாக்கியம் (Matrix sentence) என்றும், இணைக்கப்படுகின்ற வாக்கியம் உறுப்பு வாக்கியம் (Constituent sentence) என்றும் அழைக்கப்படும். எனவே கலப்பு வாக்கியம் என்பது ஒரு தலைமை வாக்கியத்தையும் ஒன்றோ அல்லது ஒன்றுக்கு மேற்பட்ட உறுப்பு வாக்கியத்தையும் கொண்டிருக்கும் வாக்கியம் ஆகும். பின்வருவன கலப்பு வாக்கியங்கள் ஆகும்

நான் இராமன் வந்தபோது போனேன்
நான் அங்கு வந்த பையனைப் பார்த்தேன்

கால வினையெச்சத் தொடர் வாக்கியம்

பின்வருவது காலவினைஎச்சத்தொடர் ஒரு முதன்மை வாக்கியத்தில் உட்பட்டு/அடங்கி வரும் ஒரு கலப்பு வாக்கியம் ஆகும்

இராமன் வந்தபோது நான் போனேன்

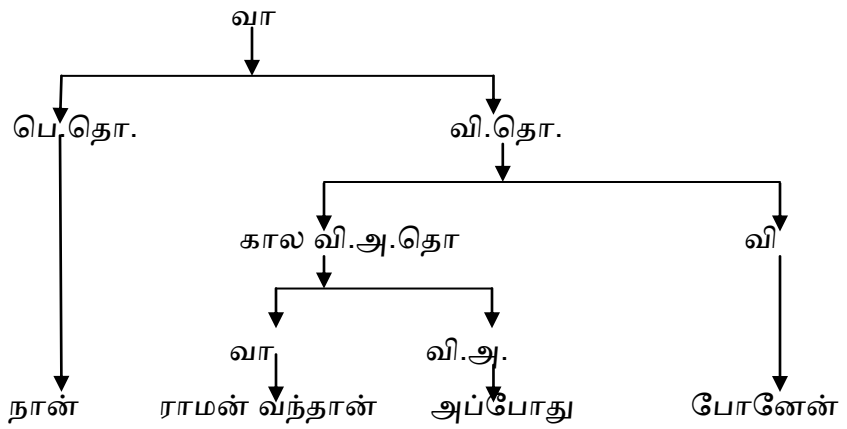
இவ்வாக்கியம் பின்னர் வரும் இரு வாக்கியங்கள் ஒன்றில் மற்றொன்று உறுப்பு வாக்கியமாக இணைவதால் ஏற்படும் ஒரு வாக்கியம் ஆகும்.

இராமன் வந்தான், அப்போது நான் போனேன்

இதில் இரண்டாவது வாக்கியம் முக்கிய அல்லது தலைமை வாக்கியமாகவும் முதல் வாக்கியம் அதில் கலந்த அல்லது அடங்கும் உறுப்பு வாக்கியமாகவும் உள்ளன. *அப்போது* என்ற வினையடையை விளக்கும் நிலையில் அமைந்துள்ளது *இராமன் வந்தான்* என்ற வாக்கியம்.

புதைநிலை (Deep structure)

இவ்வாக்கியத்தின் புதைநிலையைக் கீழ்வருமாறு காட்டலாம்.



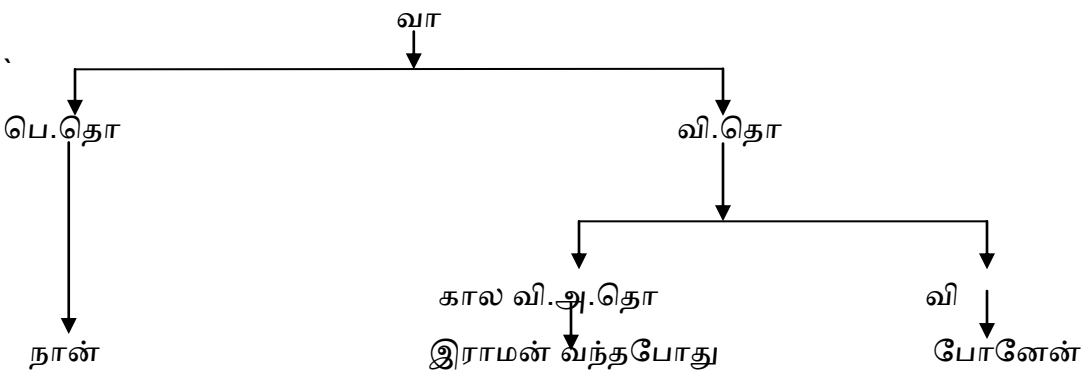
பல்வேறு மாற்றங்கள் (Transformations)

இப்புதைநிலையில் பல்வேறு மாற்றங்கள் நிகழ்கின்றன.

1. வந்தான் என்பது விகுதி. இதுவே பெயரெச்ச விகுதியாகிறது. இதனைப் பெயரெச்ச மாற்றம் (relativisation) எனக் கூறலாம்.
2. அப்போது என்பதில் உள்ள அ என்பது கெடுகின்றது. இதனை அகரச் சுட்டு கெடுதல் எனக் கூறலாம்.

புறநிலை (surface structure)

மேலே காட்டிய மாற்றங்கள் நிகழும்போது *நான் இராமன் வந்தபோது போனேன்* என்ற புறநிலை வாக்கியம் உருவாகின்றது. பின்வருமாறு இதனைக் காட்டலாம்.



பிற கலப்பு வாக்கியங்கள்

பலவிதமான கால எச்சத் தொடர்களும், வேற்றுமைத் தொடர்களும், பின்னுருபுத் தொடர்களும், வினையெச்சத் தொடர்களும், பெயரெச்சத் தொடர்களும் தலைமை வாக்கியத்தின் உறுப்புகளாக வந்து கலப்பு வாக்கியங்களை உருவாக்கும்.

உடனிகழ்ச்சி கலப்பு வாக்கியம்

பின்வரும் வாக்கியத்தில் இரு நிகழ்வுகளைக் குறிக்கும் வாக்கியங்கள் ஒன்றிற்குள் ஒன்றாக உள்ளடக்கப்பட்டு கலப்பு வாக்கியமாக வெளிப்படுகின்றன. இத்தகைய வாக்கியம் உடனிகழ்ச்சி கலப்பு வாக்கியம் எனப்படும்.

மணி பாட நான் ஆடினேன்

வேற்றுமைத் தொடர் கலப்பு வாக்கியம்

பின்வரும் வாக்கியத்தில் உட்படும் வாக்கியம் வேற்றுமைத் தொடராக முதன்மை வாகியத்தில் உட்படுத்தப்பட்டிருகின்றது. இது வேற்றுமைத் தொடர் கலப்புவாக்கியம் எனப்படும்.

நான் கண்ணன் வந்ததைக் கண்டேன்

பின்னுருபுத் தொடர் கலப்பு வாக்கியம்

பின்வரும் வாக்கியத்தில் உட்படும் வாக்கியம் பின்னுருபுத் தொடராக மற்றொன்றில் உட்படுத்தப்பட்டிருக்கின்றது. இது பின்னுருபுத்தொடர் கலப்பு வாக்கியம் எனப்படும்.

எ.கா

நான் வருவதற்கு முன்னர் அவள் வந்து விட்டாள்.

வினையெச்சத் தொடர் கலப்பு வாக்கியம்

இதில் ஒரு வாக்கியம் வினையெச்சத் தொடராக மற்றொன்றில் உட்படுத்தப்பட்டிருக்கும். இத்தகைய வாக்கியம் வினையெச்சத்தொடர் கலப்பு வாக்கியம் எனப்படும்.

எ.கா

அவள் சாப்பிட்டுவிட்டு தூங்கினாள்

பெயரெச்சத் தொடர் வாக்கியம்

இதில் ஒரு வாக்கியம் பெயரெச்சத் தொடராக மற்றொன்றில் உட்படுத்தப்பட்டிருக்கும். இத்தகைய வாக்கியம் பெயரெச்சத்தொடர் கலப்பு வாக்கியம் எனப்படும்.

எ.கா

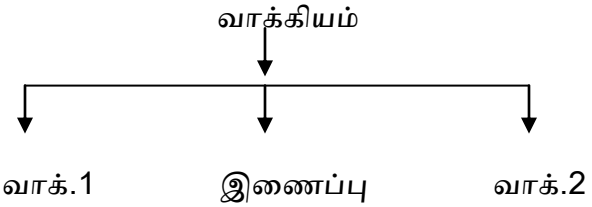
நேற்று ஊருக்குப்போன கண்ணன் இன்று பள்ளிக்கு வரவில்லை.

5.1.2.5 கூட்டு வாக்கியம்

இரண்டு வாக்கியங்கள் தலைமை வாக்கியம், உறுப்பி வாக்கியம் என்றில்லாமல் ஒத்த நிலையில் இணையும்போது உருவாகும் வாக்கியம் கூட்டு வாக்கியம் எனப்படும். பின்வருமாறு இரு வாக்கியங்களும் ஒத்தநிலையில் இணையும்போது என்ற கூட்டு வாக்கியம் உருவாகின்றது.

கண்ணன் வந்தான். (தனிநிலை வாக்கியம்)
இராமன் வந்தான். (தனிநிலை வாக்கியம்)
கண்ணனும் இராமனும் வந்தார்கள். (கூட்டுவாக்கியம்)

இதில் *கண்ணனும் இராமனும்* என்ற பெயர்த்தொடர் எழுவாயாகவும், *வந்தார்கள்* என்பது பயனிலைத் தொடராகவும் உள்ளன. இரண்டு வாக்கியங்களும் கூட்டு வாக்கியத்தில் ஒத்த நிலையில் காணப்படுகின்றன. ஒன்றின் உறுப்பாக மற்றொன்று இல்லை. இத்தகைய இணைப்பு அமைப்பைக் கீழ்க்கண்டவாறு காட்டலாம்.



கூட்டு வாக்கியங்களின் வகைப்பாடு

இரு வாக்கியங்கள் இணைந்து உருவாகும் கூட்டு வாக்கியங்கள் (1) முழுநிலை

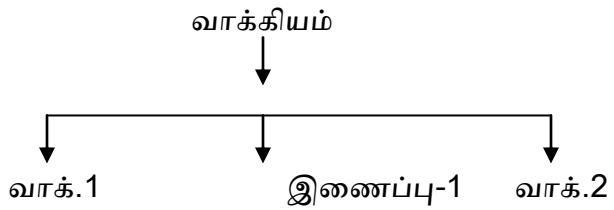
இணைப்பு வாக்கியங்கள் (conjoined sentences) என்றும், (2) பிரிநிலை இணைப்பு வாக்கியங்கள் (disjunctive) என்றும் இரண்டு வகைப்படும்.

முழுநிலை இணைப்பு

முழுநிலை இணைப்பு என்பது ஒன்றுக்கு மேற்பட்ட வாக்கியங்கள் முழு நிலையில் இணைக்கப்படுவது. அதாவது அதுவா இதுவா என்ற நிலையில் இணையாது அதுவும், இதுவும் என்ற கூட்டுநிலையில் இணைக்கப்படுவது (conjunctive).

கண்ணனும் இராமனும் வந்தார்கள்.

இதனைப் பின்வருமாறு காட்டலாம்.



பிரிநிலை இணைப்பு வாக்கியம்

இரு வாக்கியங்களை வேறுபடுத்தி அல்லது பிரித்து இணைக்கும் இணைப்பு பிரிநிலை இணைப்பு (disjunctive) எனப்படும். அதாவது 'அதுவா இல்லை இதுவா' என்ற நிலையில் இணைக்கப்படுவது

கண்ணன் வருவான் அல்லது இராமன் வருவான்.
கண்ணன் அல்லது இராமன் வருவான்.
கண்ணனோ அல்லது இராமனோ வருவார்கள்

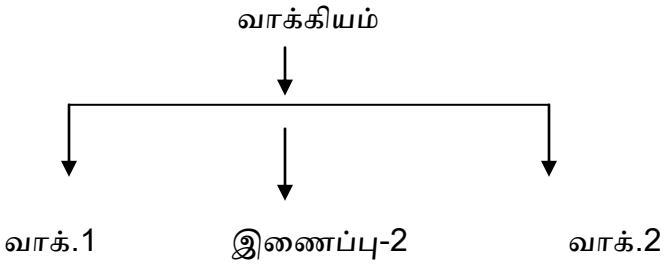
போன்ற வாக்கியங்கள் இப்பிரிவைச் சாரும்.

கண்ணன் வருவான் + இணைப்பு2 + இராமன் வருவான்

என்ற இரு வாக்கியங்களிலிருந்து உருவாவதுதான்.

கண்ணன் வருவான் அல்லது இராமன் வருவான்

போன்ற வாக்கியங்கள் கீழ்வருமாறு காட்டப்படுகிறது.



தமிழ்த்தொடரியல் அமைப்பில் ஒரு பெயர்த் தொடரும் ஒரு வினைத் தொடரும் இணைந்தோ இரு பெயர்த் தொடர்கள் இணைந்தோ ஒரு வாக்கியத்தை உருவாக்கும் வாக்கியங்கள் அவற்றின் அமைப்பின் அடிப்படையில் தனிநிலை வாக்கியம், கூட்டு வாக்கியம் மற்றும் கலவை வாக்கியம் என வேறுபடும். தனிநிலை வாக்கியத்தில் ஒரு வினைத்தொடர் காணப்படும். கூட்டு வாக்கியத்தில் ஒன்றிற்கும் மேற்பட்ட வினைத்தொடர்களோ, பெயர்த்தொடர்களோ இணைப்பான்களால் இணைக்கப்பட்டிருக்கும். கூட்டுநிலை வாக்கியத்தில் ஒன்றுக்கும் மேற்பட்ட வினைத்தொடர்கள் (அதாவது எச்சத் தொடர்கள்) வரும். பொருண்மை அடிப்படையில் வாக்கியங்களைக் கூற்று வாக்கியங்கள், கட்டளை வாக்கியங்கள், வினா வாக்கியங்கள், செய்வினை வாக்கியங்கள், செயப்பாட்டு வினை வாக்கியங்கள் எனப் பகுக்கலாம். ஒரு வாக்கியம் மற்றொரு வாக்கியத்தில் உட்படையாக வருவதன் காரணமாக வாக்கியங்கள்

கலவைத்தன்மை பெறுகின்றன. பெயரெச்சத் தொடர்கள் பெயர்த் தொடருக்குள்ளும் வினையெச்சத் தொடர்கள் வினைத் தொடருக்குள்ளும் உட்படையாக வரும்.

5.2 முந்தைய கணினித் தொடரியல் ஆய்வுகள்

முந்தைய பகுத்துக்குறித்தல் அணுகுமுறைகள் பெரும்பாலும் அகராதியால் இயக்கப்பட்டன. இதில் அகராதி உருபனியல் தரவையும், பொருண்மையியல் வர்ணனைகளையும், தொடரியல் அறிவையும் கொண்டிருக்கும். பின்னர் இலக்கண வடிவ வாதங்களின் தொடக்கம் பல வேறுபட்ட கொள்கைகள் அடிப்படையிலான பகுத்துக் குறிப்பான்களைத் தந்தது. காஷ்கெட் (Kashket,1986) என்பவர் வார்ல்பிரி என்ற மொழிக்கு சாம்ஸ்கிலின் (Chomsky) ஆளகை-கட்டுறவுக் கோட்பாடு (government and binding) அடிப்படையில் பகுத்துக்குறிப்பானை உருவாக்கினார். இவ்வணுகுமுறையில் தொடர்கள், தேர்வுக் கொள்கைகள், வேற்றுமை அடையாளப்படுத்துகை, வேற்றுமை தருகை, பங்கெடுப்பாளர் தொடர்பு படுத்துகை என்பனவற்றின் வழி உருவாக்கப்பட்டன. பங்கெடுப்பாளர் தகவல் அகராதியில் சேகரிக்கப்பட்டன. சென்குப்தா (Sengupta, et al., 1997) போன்றோர் கப்லன் மற்றும் பிரஸ்னைன் என்போரின் சொல் செயல்பாட்டு இலக்கணத்தை (Lexical Fuhnctional Grammar)வங்காள மொழிக்குத் தகுந்தவாறு மாற்றி ஒரு பகுத்துக்குறிப்பானை உருவாக்கினார். சொல் செயல்பாட்டு இலக்கணம் ஆங்கிலம் போன்ற இடம்சார் மொழிகளுக்கு (Positional languages) வலுவான வடிவவாதம் ஆகம். இது உறுப்பமைபபையும் இலக்கண உறவுகளையும் கண்டுகொள்கிறது. செங்குப்தாவின் பகுத்துக் குறிப்பான் செய்வினைப்பாடாக உள்ள எளிய வாக்கியங்களுக்கு நல்ல விளவைத் தந்தது. வினிவார்ட்டர் Winiwarter,1995)ஒன்றிணைப்பு இலக்கணத்தை (unofication grammar) விரிவாக்கி ஜொமன் மொழிக்குப் பயன்படுத்தினார். சினிவார்ட்டரின் ஒழுங்கமைப்பில் இலக்கண விதிகள் அருவமாக்கலின் பொருத்தமான மட்டத்தில் இலக்கண விதிகள் அகராதிக்கு பங்கெடுப்பாளராகச் சேர்ககப்ப்பட்டன. பெக்கா போன்றோர் (Becker et.al, 1994) V_TAG என்ற கிளை இணைப்பு இலக்கணத்திலிருந்து (TREE adjoining) மாறுபட்ட ஒன்றான பலபெயர் கால

பகுத்துக்குறித்தல் வழிமுறை வரைவை உருவாக்கினார் இப்பகுத்துக்குறிப்பானை கிளை இணைப்பு இலக்கண வடிவவாதம் முற்றிலும் உருவாக்கப்பட்ட மொழிகளுக்கு மட்டுமே பயன்படுத்த இயலும்.

பல பகுத்துக்குறிப்பான்கள் சார்புக் கொள்கைகள் (dependency principles) அடிப்படையில் உருவாக்ககப்பட்டன. கவிங்டன் (Covington, 2003) உரு சுற்று வரைபடத்தால் இயக்கப்பட்ட சொற்களுக்கு இடையே உள்ள சார்புகளை உருவாக்கி வாக்கியங்களுக்கு அமைப்பைத் தரும் ஒரு பகுத்துக்குறிப்பானை உருவாக்கினார். இதில் ஒவ்வொரு வில்லும் சொற்களின் சார்புநிலை உறவை உருப்படுத்தம் செய்யும். காரக உறவுகள் வேற்றுமைக் குறிகளுடன் பொருத்தப்படும் பாணினி இலக்கண அடிப்படையில் பாரதி போன்றோர் Gharti, et, al,1993, 2003) கட்டுப்பாடு அடிப்படையிலான பகுத்துக் குறிப்பான்களின் ஒரு குழுமத்தை உருவாக்கினார். பாரதி ஒழுங்குமுறை பகுத்துக்குறிக்கும் சிக்கலை முழுஎண் வழியமைப்புச் சிக்கலாக மொழி பெயர்த்துத் தீர்வு காண்கிறது. மூர்த்தி (Murthy,1997) இந்திய மொழிகளுக்கு உரக எச்சத்தொடர் அமைப்பு இலக்கணம் (Universal Clause Structure Grammar) என்ற இலக்கண வடிவ வாதத்தை உருவாக்கினார். இவ்வடிவவாதத்தில் மூன்று மட்ட பகுப்பாய்வு பயன்படுத்தப்படுகிறது. தொடர் அடையாளம் காணல், எச்சத்தொடர் அமைப்பு நிர்ணயம் மற்றும் பங்களிப்பு தருதல். ஒவ்வொரு மட்டத்திற்கும் அதன் பகுப்பாய்வுக்குத் தொடர்புள்ள விதிகள் இருக்கின்றன. உலக எச்சத்தொடர் அமைப்பு இலக்கணத்தில் இத்தொகுதியாக்கம் (modularization) இதைப் பிற மொழிகளுக்கு எளிதாகப் பயன்படுத்த செய்தது. சொல்நிரல் சிக்கலைக் கையாளுவதற்கான ஒரு தீர்மாங்களில் ஒன்று வரிசை மாற்று இலக்கணமாகும் (Permutational Grammar (PG)). (Geg. Ikifssib, et, al., 2001) வரிசை மாற்று இலக்கணம் அடிப்படை தொடரமைப்பு விதிகளை அவற்றின் செயல்பாட்டு உருப்படுத்தங்களுடன் குறிப்பிட்டு பின்னர் ஒரே பொருண்மையைத் தரும் பிற எல்லாத் தொடர்ச்சிகளையும் பெறவேண்டி உறுப்புகளை வரிசைமாற்றம் செய்யும். குமாரசண்முகம் மேலே குறிப்பிட்ட வேறுபட்ட அணுகுமுறைகளின் ஒரு ஒன்றிணைப்பைப் பயன்படுத்துகின்றார். அவர் தொடர்களையும் வாக்கியங்களையும் எச்சத்தொடர்களையும் கண்டுபிடிக்க மூர்த்தியின்

உலக எச்சத்தொடர் அமைப்பு இலக்கணத்தைப் போன்று நுட்பங்களைப் (Techniques) பயனபடுத்துகின்றார். சுதந்திர சொல்நிரல் சிக்கலைக் கையாள வரிசைமாற்று இலக்கணத்தைப் போன்று உறுப்புகளை வரிசை மாற்றுவதில்லை. அதற்குப் பதிலாக வாக்கியங்களின் உறுப்புகளை ஒரு குறிப்பிட்ட நிரலில் வரிசைமாற்றம் செய்கிறார். பாரதியின் ஒழுங்கமைப்பில் (Bharati, 2000) உள்ளது போன்று வாக்கியத்தின் எழுவாயையும் செயப்படுபொருளையும் கண்டுபிடிக்க வேற்றுமை உருபுகளைப் பயன்படுத்துகிறார். தமிழின் தொடரியலை விளக்க பொதுவான தொடரமைப்பு இலக்கணததைப் பயன்படுத்துகிறார்.

தமிழ் வாக்கியங்களைப் பகுதாய்வதற்கு வேண்டி ஒரு கணினியியல்சார் சட்டகத்தை உருவாக்கி பயன்படுத்துவதாகும். இதற்கு வேண்டி தமிழ்த்தொடரியலை உருப்படுத்தம் செய்ய தொடரமைப்பு விதிகள் உருவாக்கப்பட்டுள்ளன. நாற்பது மிகக் கூடுதலாக நிகழும் வினைகளுக்குப் பங்கெடுப்பாளர் அமைப்பு கண்டுபிடிக்கப் பட்டுள்ளது. பகுதாயப்பட்ட வாக்கியங்களின் தொடரியல் அமைப்புகள் பகுத்தாயப்பட்ட கிளைகளில் உருப்படுத்தம் செய்யப்பட்டுளளன. இவ்வடிவில் அவற்றை இயந்திர மொழிபெயர்ப்பு மற்றும் தகவல் பிரித்தெடுப்பு இவற்றை உள்ளடக்கிய பயன்பாடுகளில் நேரடியாகப் பயன்படுத்த இயலும்.

உள்ளீட்டுத் தமிழ் வாக்கியத்தை முக்கியப் பகுப்பான் பகுப்பதற்கு முன் அது பகுப்பாய்வின் பல மட்டங்களைக் கடந்துசெல்ல வேண்டும். கீழே தந்துள்ள படத்தில் செவ்வகங்கள் பகுப்பாய்வு செய்யும் பகுப்பாய்வுத் தொகுதிகளையும் (analyzer modules) வட்டங்கள் அவை பயன்படுத்தும் பொருத்தமான மூலப்பொருள்களையும் உருப்படுத்தம் செய்யும்.

தமிழுக்கு வேண்டி சில பெயர்த்தொடர் பகுப்பான்கள் உருவாக்கப்பட்டுள்ளன. அவை கீழே விளக்கப்பட்டுள்ளன.

5.2.1 AUKBCRC -இன் தமிழுக்கான பெயர்த்தொடர் பகுப்பான்

இவ்வணுகுமுறை விதி அடிப்படையிலானது. இந்நெறிமுறையில் ஒரு தரவுத்தொகுதி எடுக்கப்பட்டு இரண்டோ அதற்கு அதிகமான குமங்களாகப் பிரிக்கப்படும். இப்பிரிக்கப்பட்ட குழுமத்தில் ஒன்று பயிற்சித் தரவாகப் பயன்படுத்தப்படும். பயிற்சித் தரவுக் குழுமம் எடுக்கப்பட்டு பெயர்த்தொடருக்காக மனித முயற்சியால் பகுக்கப்படும். இவ்வாறு ஒரு வாக்கியத்தில் உள்ள பெயர்த் தொடர்களைப் பிரிப்பதற்குப் பயன்படுத்தப்படும் விதிகள் உருவாக்கப்படும். இவ்விதிகள் தொடர் பகுப்பிற்கு அடிப்படையாக அமையும். தொடர் பகுப்பான் வழியமைப்பு இவ்விதிகளைப் பயன்படுத்தி பரிசோதனைத் தரவை பகுக்கும். இவ்விதிகளின் பயன்பாட்டுப் பரப்பு இத்தரவுக்குழுமத்தில் பரிசோதிக்கப்படும். துல்லியமும் மீள அழைத்தலும் இதற்குக் கணிக்கப்படும். விளைவு ஒழுங்குமுறையின் பயன்பாட்டுப் பரப்பை விருத்தி செய்ய கூடதல் விதிகள் தேவையா என்று பரிசோதிக்கப்படும். பின்னர் ஒழுங்குமுறை வேறுபட்ட பிற பயன்பாடுகளுக்குப் பரிசோதிக்கப்படும்.

5.2.2 RCILTS –இன் வானவில்

வானவில் இந்திய மொழிகளின் தொழில்நுட்ப முன்னேற்றம் என்ற திட்டத்தின் கீழ் தமிழுக்காகச் செய்யப்பட்ட இயற்கைமொழி ஆய்வுகளில் ஒன்றாகும். இது ஒரு தமிழ் வாக்கியத்தின் தொடரியல் உறுப்புகளை அடையாளம் காண்கிறது. இது பட்டியல் வடிவில் பகுப்புக் கிளைப் படத்தை (parse tree) வெளிப்யிடுகின்றது. இது தனிநிலை வாக்கியங்களைய கூட்டு வாக்கியங்களையும் கையாளுகின்றது. தனிநிலை வாக்கியங்கள் ஒரு வினையையும் பல பெயர்த் தொடர்களையும், தனிநிலை வினையடைகளையும், பெயரடைகளையும் கொண்டிருக்கும். பல எச்சத் தொடர்கள் உள்ள வாக்கியங்களின் தேர்வுகளில் வானவில் துப்புச்சொற்கள் மற்றும் தொடர்கள் அடிப்படையில் எச்சத் தொடர்களைத் தொடர் குழுக்களாகக் குழுமும். இது தொடரியல் அமைப்பு

இலக்கணத்தை உருவாக்கும். வானவில் சொல்நிரலைக் கையாள முன்நோக்குதலைப் பயன்படுத்தும்.

பகுத்துக்குறித்தல் என்பது ஒரு வாக்கியத்தை அறிந்துகொண்டு அதற்குத் தொடரியல் அமைப்பைத் தருவதாகும். இது தொடரியல் ஆய்வில் மிகப்பெரிய சிக்கலாகும். இதன் காரணமாக இது இயற்கை மொழி ஆய்வின் ஒரு பெரும்பகுதியாக அமைகிறது.

நிர்ணயிக்கப்பட்ட பகுப்பானின் உள்ளீடு (input) அடையாளப்படுத்தப்பட்ட சொற்களின் நிரலான தொடர்ச்சியும் தொடரமைப்பு விதிகளின் ஒரு குழுமமும் ஆகும். இதன் வெளியீடு (out put) வாக்கியத்தின் பகுக்கப்பட்ட அமைப்பாகும். ஒரு வாக்கியத்தின் பகுக்கப்பட்ட அமைப்பு அதன் எழுவாய், செயப்படுபொருள் மற்றும் பயனிலைகளை அறிந்துகொள்ள உதவுகிறது. மேலும், பகுக்கப்பட்ட கிளை அமைப்பு பிற தொடர்களையும் அடை செய்யும் தொடர்களையும் அடையாளம் காண்கிறது.

ஒரு மொழிக்கு ஒரு மாதிரி பகுப்பான் ஒரு அடையாளப்படுத்தியைப் பயன்படுத்தி வாக்கியங்களின் சொல் வகையை அடையாளப்படுத்துகிறது. பின்னர் பகுக்கும் வழிமுறை வரைவு (algorithm) வாக்கியத்தைப் பகுப்பாய்வு செய்து அம்மொழி இலக்கணத்திற்குத் தரப்பட்ட தொடரமைப்பு விதிகிளின் அடிப்படையில் பொருத்தமான அமைப்பைத் தரும். பகுப்பாய்வின் போது வழிமுறை வரைவு உறுப்புகளுக்கு இடையேயுள்ள இலக்கண உடன்பாட்டையும் முக்கிய வினைக்கம் செயப்படுபொருள் தொடர்களுக்கும் இடையே உள்ள சார்பு உறவையும் பரிசோதிக்கும்.

5.2.3 குமாரசண்முகத்தின் தொடரியல் பகுப்பாய்வு

குமார சண்முகம் தமது எம்.எஸ். பட்டப்படிப்புக்காக தமிழில் தொடரியல் பகுப்பாய்வு என்ற ஆய்வேட்டைச் சமர்ப்பித்துள்ளார். அதன் சுருக்கம் இங்கு தரப்பட்டுள்ளது.

5.2.3.1 வாக்கியங்களின் இலக்கணப் பண்புக்கூறுகள்

வாக்கியத்தில் உள்ள சொற்களின் சொல் வகைப்பாடும் பண்புக்கூறு தகவலும் வாக்கியத்தின் பகுப்பாய்வுக்கு முக்கியமாகும். பின்வரும் அட்டவணை சொல் வகைப்பாடுகளையும் அவற்றுடன் தொடர்புடைய அடையாளங்களையும் அவற்றின் பண்புக்கூறுகளுடன் காட்டும்.

சொல் வகைப்பாட்டுடன் தொடர்புடைய அடையாளப்படுத்திகள்

வ.எண்	சொல் வகைப்பாடு	அடையாளம்/உருப்படுத்தம்	எடுத்துக்காட்டு
1.	முற்றுவினை	V	பார்த்தான்
2.	பெயர்	N	பையன்
3.	மாற்றுப்பெயர்	P	அவர்
4.	பெயரடை	J	நல்ல
5.	வினையடை	R	வேகமாக
6.	பின்னுருபு	I	முன்னால்
7.	சுட்டடை	D	அந்த
8.	அளவடை	G	பல
9.	பெயர் மாறுபடுத்தி	M	புதுமை
10.	கிழமைப்பெயர்	GN	பொருளின்
11.	முற்றுப்பெறாத வினை	NV	பெற்று

12.	பெயரெச்ச வினை	RV	படித்த
13.	நிரப்பிய வினை	CV	என்று
14.	செயல்பாட்டுச் சொல்	F	வரை

5.2.3.2 இலக்கணப் பண்புகளும் அவற்றுடன் தொடர்புடைய அடையாளப்படுத்திகளும்

வ.எண்	பண்புகள்	சொல்வரை/உருப்படுத்தம்
பால்		
1.	ஆண்பால்	+M, -F
2.	பெண்பால்	-M, +F
3.	பலர்பால்	+M, +F
4.	ஒன்றன் பால்	-M, -F
எண்		
5.	ஒருமை	+S
6.	பன்மை	-S
இடம்		
7.	தன்மை	+1, -2
8.	முன்னிலை	-1, +2
9.	படர்க்கை	+1, +2
காலம்		
10.	நிகழ்காலம்	+P, -A

11.	இறந்தகாலம்	-P, +A
12.	எதிர்காலம்	+P, +A
வேற்றுமை		
13.	செயப்படுபொருள் வேற்றுமை	+ACC
14.	கொடை வேற்றுமை	+DAT
15.	இட வேற்றுமை	+LOC
16.	கருவி வேற்றுமை	+INS
17.	உடனிகழ்ச்சி வேற்றுமை	+SOC
18.	நீங்கல் வேற்றுமை	+ABL
19.	பயன் வேற்றுமை	+BEN
20.	கிழமை வேற்றுமை	+GEN

5.2.3.3 உருபனியல் பகுப்பாய்வு

உருபனியல் பகுப்பாய்வு சொற்களைப் பகுக்கும் செயல்பாட்டையும் அவற்றின் ஒட்டு உருபன்பளையும் வேர்களைக் கண்டுபிடிக்கும் செயல்பாட்டையும் உள்ளடக்கும். இவ்வுருபனியல் பகுப்பாய்வினால் ஒரு சொல்லின் சொல் வகைப்பாட்டையும் அதன் இலக்கணப் பண்புக்கூறுகளையும் அடையாளம் காண இயலும்.

5.2.3.4 சொல் வடிவமும் அதன் அடையாளங்களும்

சொல்வடிவம்	**உருபன் விடுவரல்**	**சொல்லும் அதன் அடையாளப்படுத்தியும்**

கேட்டான்	கேள்<verb>+த்த்<past tense>+ஆன்<3rd person, singular,masculine>	கேட்டான் <V:+p,-A, +1.+2 +3, +S, +M, -F>
மரத்தை	மரம் <noun> + ஐ <accusative case>	மரத்தை <N:+ACC, +S, -M, -F>

5.2.3.5 சொல்வகை அடையாளப்படுத்தல்

சொல்வகை அடையாளப்படுத்தல் வாக்கியங்களின் உறுப்புகளுக்குச் சரியான அடையாளத்தைத் தரும் செயல்பாடாகும். தமிழ் போன்ற மொழிகளுக்கு உருபனியல் பகுப்பாய்வி 70% துல்லியத்துடன் சொல்வகைப்பாட்டை கண்டுபிடிக்க இயலும். இருப்பினும் பல நிகழ்வுகளில் உருபனியல் பகுப்பாய்வி ஒரு சொல்லுக்கு ஒரு அடையாளத்தை மயக்கமின்றி தர இயலாது. இம்மாதிரியான மயக்கங்களை நீக்கக் குமாரசண்முகம் உருபனியல் பகுப்பாய்விற்குப் பின்னர் சொல்வகை அடையாளப்படுத்தியைப் பயன்படுத்துகிறார். சொல்வகை அடையாளப்படுத்தி உருபனியல் ஆய்வியின் விடுவரல் (output) ஏற்று இயலுகின்ற அளவு சொல்மட்ட விதிகளைப் பயன்படுத்தி மயக்கத்தை நீக்குகிறது. எஞ்சுவது பகுத்துக்குறிப்பானால் கவனித்துக் கொள்ளப்படும்.

5.2.3.6 இயல்பாக்கம் (Normalization)

தமிழ் ஒத்தறி அடிப்படையில் சொல்நிரல் சுதந்திரமுள்ள மொழியாகையால் இறுதி வினைக்கு முன்னர் வரும் பெ.தொ பங்கெடுப்பாளர்கள் இறுதி வினைக்கு முன் எந்த வரிசையிலும் வந்து ஒரே பொருளைத் தரும்.

எ.கா

ஆசிரியர் குழந்தைக்குப் பரிசைக் கொடுத்தார்.

ஆசிரியர் பரிசைக் குழந்தைக்குக் கொடுத்தார்.

பரிசை ஆசிரியர் குழந்தைக்குக் கொடுத்தார்.

பரிசைக் குழந்தைக்கு ஆசிரியர் கொடுத்தார்.

குழந்தைக்குப் பரிசை ஆசிரியர் கொடுத்தார்.

குழந்தைக்கு ஆசிரியர் பரிசைக் கொடுத்தார்.

குமாரசண்முகம் இயல்பாக்கத்திற்காக வாக்கியங்களில் தொடர்கள் ஒரு அனுமதிக்கப்பட்ட எல்லைக்குள் ஏற்கெனவே தீர்மானிக்கப்பட்ட வரிசைகளில் வருவதாக எடுத்துக் கொள்கிறார். இயல்பாக்கி வாக்கியத்தை வரிசையாக ஆய்ந்து தொடர் எல்லைகளை அடையாளம் காண்கின்றது. அதைத் தொடர்ந்து அது எச்சத் தொடர்களையும் பெயர்த்தொடர்களையும் இறுதி வினைக்கு முன் ஒரு குறிப்பிட்ட நிலையில் வரிசைப்படுத்துகிறது. பெயர்த்தொடர்களின் மற்றும் எச்சத் தொடர்களின் எல்லைகள் கண்டுபிடிக்கப்பட்ட பின் இயல்பாக்கி இவற்றை வலமிருந்து இடமாக பின்வரும் முன்நிர்ணயிக்கப்பட்ட நிரலில் மறுவரிசைப்படுத்தும்.

0. வினை,முற்றுப்பெறா வினை,பெயரெச்ச வினை,நிரப்பு வினை

1. பெ.தொ+செயப்படுபொருள் வேற்றுமை

2. பெ.தொ+கொடை வேற்றுமை

3. பின்னுருபு+இட வேற்றுமை

4. பின்னுருபு+கருவி வேற்றுமை

5. பின்னுருபு+உடனிகழ்ச்சி வேற்றுமை

6. பின்னுருபு+நீங்கல் வேற்றுமை

7. பின்னுருபு+பயன் வேற்றுமை

8. வினையடை+கிழமை வேற்றுமை

9. எழுவாய்

எடுத்துக்காட்டாக முன்னர் பட்டியலிடப்பட்ட வாக்கியங்களில் எது தரப்பட்டாலும் இயல்பாக்கி பின்வரும் வெளிப்யீட்டைத் தரும்.

ஆசிரியர் குழந்தைகளுக்குப் பரிசைக் கொடுத்தார்.

5.2.3.7 இலக்கணச்சட்டகம்

பகுப்பாய்வுகுக் குமாரசண்முகம் தொடரமைப்பு இலக்கணத்தைப் பின்பற்றுகிறார். சொல் வகைப்பாடுகளில் இருந்து கிளைகளை உருவாக்க X-பார் கோட்பாட்டை ஒக்கும் அணுகுமுறையைப் பின்பற்றுகிறார்.

தொடரமைப்பு விதிகள் ஒரு குழும விதிகளைக் கொண்டது. சண்முகசுந்தரம் பின்பற்றுகின்ற விதிகள் கீழே பட்டியலிடப்பட்டுள்ளன.

Sl.No	Rules	Examples
	Quantifier Phrase Rules	**Q->Quantifier**
1	Q '=> Q	Q(Q(சில)) Q(Q(few))
2	QP => Q	QP (Q(சில)) QP (Q(few))
	Adjectival Pharase	**j->Adjective**

	(ADJP) Rules	
3	J => J	J'(J(நல்ல)) J'(J(good))
4	ADJP => J	ADJP(நல்ல)) ADJP(J'(good))
	Modifier Phrase (MP) Rules	**M->Noun modifier**
5	M => M	M'(M'(கலை)) M'(M'(art))
6	MP => M	MP'(MP'(கலை)) MP (MP'(art))
	Genitive Phrase (GP) Rules	**GN->Genitive noun D->Determiner** **RC->Relative Clause**
7	GN => *GN	GN'(GN'(மாணவனுடைய)) GN'(GN'(student's))
8	N => MP*GN	GN'(MP(கல்லூரி) GN'(மாணவனுடைய)) GN'(MP'(college)GN'(student's))
9	GN => ADJP*GN	GN'(ADJP (நல்ல) GN'(மாணவனுடைய)) GN'(ADJP(good) GN'(student's))
10	GN'=> RC*GN	GN'(RC(<trace>படத்தைப் பார்த்த) GN'(பையனுடைய)) RC(<trace>movie saw+RP)GN'(boy's))
11	GN'=>	GN'GP(மாணவனுடைய)GN'(புத்தகத்தின்))

	GP*GN	GN’(GP(student’s) GN’(book’s))
12	GP=>*GN	GN’(GP’ மாணவனுடைய)) GP’(GP’(student’s))
13	GP=>D*GN ’	GP(D(அந்த)GN’(மாணவனுடைய)) GP(D(that)GN’(student’s))
	Noun Phrase (NP) Rules	**P-> Pronoun N-> Noun** **F-> Functional word**
14	N’ => *N	N’(N(கல்லூரி)) N’(N(college))
15	N’ => P *N	N’(P(அவன்) N’(P(he))
16	N’ => MP *N’	N’(MP(fiy)N’(கல்லூரி)) N’(MP(art)N’(collegep))
17	N’ => QP *N’	N’(QP(சில)N’(கல்லூரி)) N’(QP(few)N’(college))
18	N’ => ADJP*N’	N’(ADJP(நல்ல)N’(கல்லூரி)) N’(ADJP(good)N’(college))
19	N’ => RC*N’	N’(RC(<tarace> படத்தைப் பார்த்த)N’(பையன்)) N’(RC(<tarace>movie saw+RP)N’(boy))
20	N’ => GP*N’	N’(GP(மாணவனுடைய)N’(புத்தகம்)) N’(GP(student’s)N’(book))
21	N’ => N’F	N’(N’(வீடு)F(வரை)) N’(N’(house)F(till))
22	NP => *N’	N’(N’(கல்லூரி)) N’(N’(college))

23	NP => D*N’	NP(D(அந்த)N’(கல்லூரி)) NP(D(that)N’(college))
	Postpositio n Phrase (PP) Rules	**I -> Postposition** **N’ -> N’wutg itger tgan Accusative and Dative cases**
24	PP => *N’	PP(N’(கல்லூரியில்)) PP(N’(college in))
25	PP => D*N’	PP(D(அந்த)N’(கல்லூரியில்)) PP(D(that)N’(college in))
26	P => *N’I	PP(N’(வீட்டிற்கு)I(முன்)) (N’(house in)I(front of))
	Adverbial Pharase (ADVP) Rules	**R -> Adverb**
27	R’ => R	R’(R(மெதுவாக)) R’(R(slowly))
28	R’=> QPR’	R’(QP(மிக)R’(மெதுவாக)) R’(QP(very)R’(slowly))
29	ADVP => R’	ADVP(R’(மெதுவாக)) ADVP(R’(slowly))
	Verrb Phrase (VP) Rules	**V -> Verb**
30	V’ => *V	V’(V(பார்த்தான்))

		V'(V(saw))
31	V' => NP *V	V'PP(காடுகளில்)V(காணப்படுகிறது)) V'PP(in forest)V(seen))
32	V' => PP*V	V'(NP(படத்தைப்)V(பார்த்தான்)) V'(NP(movie)V(saw))
33	V' => PP PP *V	V'(ஜோடியுடன்PP(காடுகளில் V(காணப்படுகிறது)) V'(with pair)PP(in forest)V(seen))
34	V' => NP NP *V	V'(NP(அவனுக்கு)NP(புத்தகத்தைக்)V(கொடுத்தேன்)) V'(NP(him)NP(book)V(gave))
35	V' => PP NP *V	V'(PP(திரையரங்கில்)NP(படத்தைப்)V(பார்த்தான்)) V'(PP(in theatre)NP(movie)V(saw))
36	V' => PP PP NP *V	V'(PP(திரையரங்கில்)PP(நண்பருடன்)NP((படத்தைப்)V(பார்த்தான்)) V'(PP(in theatre)PP(with friend)NP(movie)V(saw))
37	V' => PP NP NP *V	V'(PP(வீட்டில்)NP(அவனுக்கு)NP(புத்தகத்தைக்)V(கொடுத்தேன்)) V'(PP(at home)NP(him)NP(book)V(gave))
38	V' => NC *V'	V'(NC(படத்தைப் பார்த்து)V'(கருத்தைச் சொன்னான்)) V'(NC(movie saw+VBP)V'(comments said))
39	V' => ADVP *V'	V'(ADVP(கவனமாகப்V'(பார்த்தான்)) V'(ADVP(carfully)V'(saw))
40	V' => *V'	VP(V'(பார்த்தான்)) VP(V'(saw))
	Nonfinite Verb Phrase (NVP)	**NV -> Nonfinite Verb NC->Non-finite clause VBP-> Verbal Participial**

	Rules	
41	NV’ => *NV	NV’(NV(பார்த்து)) NV’(NV(saw))
42	NV’ => PP *NV	NV’(PP(காடுகளில்NV(காணப்பட்டு)) NV’(PP(in forest)NV(seen+VBP))
43	NV’ => NP *NV	NV’(NP(படத்தைப்)NV(பார்த்து)) NV’(NP(movie)NV(saw+VBP))
44	NV’ => PP PP *NV	NV’(PP(ஜோடியுடன்)PP(காடுகளில்)NV(காணப்பட்டு)) NV’(PP(with pair)PP(in forest)NV(seen+VBP))
45	NV’ => NP NP *NV	NV’(NP அவனுக்கு)NP(புத்தகத்தைக்)NV(கொடுத்து)) NV’(NP(him)NP(book)NV(gave+VBP))
46	NV’ => PP NP *NV	NV’(PP(திரையரங்கில்)PP(படத்தைப்)NV(பார்த்து)) NV’(PP(in theatre)PP(movie)NV(saw+VBP))
47	NV’ => PP PP NP *NV	NV’(PP(திரையரங்கில்)PP(நண்பனுடன்PP(படத்தைப்)NV(பார்த்து)) NV’(PP(in theatre)P P(withfriend)PP(movie)NV(saw+VBP))
48	NV’ => PP NP NP *NV	NV’(PP(வீட்டில்NP(அவனுக்கு)NP(புத்தகத்தைக்)NV(கொடுத்து)) NV’(PP(at home)NP(him)NP(book)NV(gave+BVP))
49	NV’ => NC *NV’	NV’(NC(படத்தைப் பார்த்து)NV’(கருத்தைச் சொல்லி)) NV’(NC(movie saw+VBP)NV’(comments said+VBP))
50	NV’ => ADVP *NV’	NV’(ADVP(கவனமாகப்)NV’(பார்த்து)) NV’(ADVP(carefully)NV’(saw+VBP))
51	NV’ => *NV’	NVP(NV’(பார்த்து)) NVP(NV’(saw_VBP))
	Nonfinite Clause	CV->Complementizing Verb

	(NC) Rules	
52	NC => *NP *NVP’	NC’(NP(அவன்)NVP(படத்தைப் பார்த்து)) NC’(NP(he)NVP(movie saw+VBP))
53	NVP => *NV’	NC(S(அவன் படத்தைப் பார்த்தான்)CVஎன்று)) NC(S(he movie saw)CV(that))
	Relativizing Verb (RVP) Rules	**RV -> Relativizing Berb** **RP -> Relative Participle**
54	RV’ => *RV	RV’(RV(பார்த்த)) RV’(RV(saw+RP))
55	RV’ => PP *RV	RV’(PP(காடுகளில்)RV(காணப்பட்ட)) RV’(PP(in forest)RV(seen+RP))
56	RV’ => NP *RV	RV’(NP(படத்தைப்RV(பார்த்த)) RV’(NP(movie)RV(saw+RP))
57	RV’ => PP PP *RV	RV’(PP(ஜோடியுடன்)PP(காடுகளில்)RV(காணப்பட்ட)) RV’(PP(with pair)PP(in forest)RV(seen+RP))
58	RV’ => NP NP *RV	RV’(NP(அவனுக்கு)NP(புத்தகத்தைக்)RV(கொடுத்த)) RV’(NP(him)NP(book)RV(gave+RP))
59	RV’ => P NPP *RV	RV’(PP(திரையரங்கில்)PP(படத்தைப்)RV(பார்த்த)) RV’(PP(in theatre)PP(movie)RV(saw+RP))
60	RV’ => PP PP NP *RV	RV’(PP(திரையரங்கில்)PP(நண்பனுடன்)PP(படத்தைப்)RV(பார்த்த)) RV’(PP(in theatre)P P(withfriend)PP(movie)RV(saw+RP))
61	RV’ => PP NP NP *RV	RV’(PP(வீட்டில்)NP(அவனுக்கு)NP(புத்தகத்தைக்)RV(கொடுத்த)) RV’(PP(at home)NP(him)NP(book)RV(gave+RP))
62	RV’ => NC	RV’(NC(படத்தைப் பார்த்து)RV’(கருத்தைச் சொன்ன))

	*RV’	RV’(NC(movie saw+VBP)RV’(comments said+RP))
63	RV’ => ADVP *RV’	RV’(ADVP(கவனமாகப்RV’(பார்த்த)) RV’(ADVP(carefully)RV’(saw+RP))
64	RV’ => *RV’	RVP(RV’(பார்த்த)) RVP(RV’(saw+RP))
	Relative Clause (RC) Rules	
65	RC => *NP *VP	RC(NP(<trace>)RVP(படத்தைப் பார்த்த)) RC(NP(<trace>)RVP(movie saw+P))
	Sentence (S) Rules	
66	S => *NP *VP	S(NP(mtd)VP(அவன் படத்தைப் பார்த்தான்)) S(NP(he)VP(movie saw))

ஒரு விதியின் தலைப்பு உறுப்பு ‘என்ற குறியீட்டால் முன் தொடரப்படும் வாக்கியம் பகுத்துக் குறிக்கப்படும்போது தலைமை உறுப்பின் இலக்கண பண்புக்கூறுகள் இடப்பக்கமுள்ள இறுதியுறாத ஒன்றிற்குக் கடத்தப்படுகிறது. M என்ற சொல் வகைப்பாடு கூட்டுப்பெயர்களைக் கையாளவேண்டி பயன்படுத்தப்படுகிற தனித்தன்மையான வகைப்பாடாகும். ‘M’ என்பது மாறுபடுத்தும் தொடரியல் (Modifier Phrase (MP)) தன்மை உறுப்பாகும். கிழமை வேற்றுமைக் குறியீடு உள்ள ஒரு பெயர் மட்டும்தான் கிழமைத்தொடரின் (Genitive Phrase/GP) தலைமையாக வரும். பின்னுருபுத் தொடர்கள் பெயரைத் தொடர்ந்து முன்,பின் என்ற பின்னுருபுகளால் மெய்ப்படுத்தம் செய்யப்படும். முற்றுப்பெறாத எச்சத்தொடரில் (Non-finite Clause/NC) தலைமை ஒரு முற்றுப்பெறாத வினையாகும். தமிழில் நான்கு வகையான முற்றுப்பெறா வினைகளிலிருந்து நான்கு

வகையான முற்றுப்பெறா எச்சத்தொடர்கள் கிடைக்கின்றன. எச்சத்தொடர் (Relative clause (RC)) பெயர்களுக்கு முன்னர் வந்து அவற்றை அடை செய்யும்.

6. இந்தியமொழியிலிருந்து இந்திய மொழி மொழிபெயர்ப்பு ஒழுங்கு முறையில் தொடர் பகுப்பாய்வு

இந்தியமொழியிலிருந்து இந்திய மொழி மொழிபெயர்ப்பு ஒழுங்கு முறையில் தொடர் பகுப்பாய்வு (chunking) ஒருவாக்கியத்தின் தனிநிலை பெயர்தொடர்கள், வினைக் குழுமங்கள், பெயரடைத் தொடர்கள், வினையடைத் தொடர்கள் இவற்றை அடையாளங்காணுதலைக் குறிப்பிடும். இது தொடர்களின் எல்லையையும் புலக்குறிப்பையும் கண்டுபிடிப்பதை உட்படுத்தும்.

இடுவரல்-விடுவரல் தனிக்குறிப்பீடுகள்

Input-Output Specification

Input: TKN_, CAT_

Output: ADDR_ (ADDR என்பதன் மதிப்பு மாறுகையில் மீளமைப்புச் செயல்பாடு நடக்கிறது என்பது இதன் அர்த்தமாகும்.)

மேலே தரப்பட்டுள்ள இடுவரல் தனிக்குறிப்பீடுகள் SSF-இல் வருணனை செய்வதற்கு TKN_ மற்றும் CAT_ என்ற தனிப்பண்புகள் அதற்குத் தேவை அல்லது அது பன்படுத்துகின்றது என்று அர்த்தம்தரும். வேறுவிதமாகக் கூறினால் இடுவரல் வாக்கியத்தில் சொற்கள் அல்லது டொக்கன்கள் சொல்வகைப்பாட்டு வகையுடன் தரப்பட்டுள்ளன.

மேலே தரப்பட்டுள்ள விடுவரல் தனிக்குறிப்பீடுகள் ‘ADDR_’ இந்தத் தொகுதியால் வரையறவிளக்கம் பெறும் என்று அர்த்தம் பெறும்.

எடுத்துக்காட்டு:

வருணனை: தனிநிலைப் பெயர்த்தொடர்கள் தொடர் உறுப்புகளாகக் (NPs) குழுமப்படும். வினைத் தொடர்ச்சிகள் வினைத் தொடர் உறுப்புகளாகக் (VG) குழுமப்படும்.

6.1 இ-இ இயந்திர மொழிபெயர்ப்புத் திட்டத்திற்குத் தேர்ந்தெடுக்கப்பட்டுள்ள தொடர்கள்

இதில் தொடர்களின் வரையறை விளக்கமும் தொடர் வகைகளும் சில தனித்துவம் வாய்ந்த நேர்வுகளும் பேசப்படும். இந்திய மொழிகளுக்கு இடையிலான மொழிபெயர்ப்புத் திட்டத்தில் 11 தொடர்கள் தெரிந்தெடுக்கப்படு அவை தொடர்களுக்காக அடையாளப் படுத்தலுக்குப் பயன் படுத்தபடுகின்றன.

முக்கிய தொடர்களின் (chunks) குழுமம் அவற்றின் உருப்படுத்தமும்

	முக்கிய தொடர்களின் குழுமம்	**உருப்படுத்தம்**
1	பெயர்த் தொடர்	NP
2	வினைமுற்றுத் தொடர்	VGF
3	வினைமுற்றாத் தொடர்	VGNF
4	வினைப்பெயர்த் தொடர்	VGNN
5	வினை எச்சத்தொடர்	VGINF
6	பெயரைடைத் தொடர்	JJP
7	வினையடைத் தொடர்	RBP
8	இணைப்புத் தொடர்	CCP
9	வியப்புத் தொடர்	INTJ
10	எதிர்மறைத் தொடர்	NEGP
11	வெற்றிடத் தொடர்	BLK

6.1.1 பெயர்த் தொடர் (NP)

பெயர்த்தொடர்கள் (NP chunks) NP என அடையாளப் படுத்தப்படும்; இது மறுதரவாக வராத பெயர்த்தொடர்களையும் பின்னுருபுத் தொடர்களையும் உட்படுத்தும். பெயர்த்தொடரின் தலைமை ஒரு பெயராகும். சிறப்புசெய்யும் சொற்கள் பெயர்த்தொடரின் இடது பக்க எல்லையை நிறுவும்; வேற்றுமையோ தலைமைப் பெயரோ வலது பக்க எல்லையை நிறுவும். பெயரை அடைசெய்யும் வருணனைப் பெயரடை(கள்) பெயர்த்தொடர் உறுப்பின் பகுதியாக அமையும். தலைமைப் பெயருடன் நங்கூரமிட்டுள்ள குறைச் சொற்களும் பெயர்த்தொடருடன் குழுமப்படும். அது பெயருக்குப் பின்னரோ வேற்றுமைக்குப் பின்னரோ வந்தால், அது தொடரின் வலது பக்க எல்லையை நிர்ணயிக்கும். பெயர்த்தொடர்களின் சில எடுத்துக்காட்டுகள் கீழே தரப்பட்டுள்ளன.

((குழந்தைகள்_NN))_NP, ((சில_QF குழந்தைகள்))_NP, ((சில_QF நல்ல_JJ குழந்தைகள்_NN))_NP, ((மேசை மேல்_PSP))_NP, ((ஒரு_QC கறுத்த_JJ குதிரை_NN))_NP, ((இந்த_DEM புதிய_JJ புத்தகம்_NN))_NP, ((இந்த_DEM புதிய_JJ புத்தகத்தின்_NN மேல்_PSP))_NP

6.1.2 வினைத் தொடர் (VG)

வினைத் தொடர் (verb chunks) பலவகைப்படும். ஒரு வினைக் குழுமம் முக்கிய வினையையும் இருப்பின் துணை வினைகளையும் உள்ளடக்கும். பின் வருவது வினைத் தொடர் உறுப்புகளின் எடுத்துக்காட்டுகளாகும்.

((போனான்)), ((போய்விட்டான்)), (போய்க்கொண்டிருக்கிறான்))

6.1.3 பெயரைடைத் தொடர் (JJP)

பெயரடைத் தொடர் JJP என அடையாளப்படுத்தப்படும். இத்தொடர் பயனிலைப் பெயர்தொடர்களையும் உள்ளடக்கி எல்லாப் பெயர்த்தொடர் உறுப்புகளையும் கொண்டிருக்கும். இருப்பினும் பெயருக்கு முன்னர் வரும் பெயரடைகள் பெயர்த் தொடருடன் சேர்த்துக் குழுமப்படும்.

6.1..4 வினையடைத் தொடர் (RBP)

இத்தொடர் உறுப்பின் பெயரும் சொல்வகை அடையாளப் படுத்தலின் போது பயன்படுத்தப்படுகிற சொல்வகை அடையாளத்தின் அடிப்படையில் அமைகின்றது. எல்லா வினையடைத்தொடர் RBP என அடையாளப்படுத்தப்படும். இந்தத் தொடர் எல்லாச் சுத்தமான வினைத்தொடர்களையும் உட்படுத்தும்.

6.1.5 இணைப்புத் தொடர் (CCP)

இணைப்புத் தொடர் உறுப்புகள் யாவும் CCP என அடையாளப்படுத்தப்படும். இத்தகவல் பெரிய வாக்கிய அமைபுகளை உருவாக்கத் தேவையாகும்.

6.1.6 வியப்புத் தொடர் (INTJ)

எல்லா வியப்புத் தொடர்களும் INTJ என அடையாளப்படுத்தப்படும்.

6.1.7 எதிர்மறைத் தொடர் (NEGP)

வினையைச் சுற்றி இருக்கும் எதிர்மறைக் குறைச்சொல் வினைக் குழுவுடன் குழுமப்படவேண்டும். அது வினையிலிருந்தோ பெயரிலிருந்தோ விலகி இருந்தால் அது தனியான தொடராக குழுமப்படவேண்டும்.

5.2. தொடர் வரையறுத்தலின் எடுத்துக்காட்டுகள்

AU-KBC-நிறுவனதில் நடைபெற்றுவரும் இந்திய மொழியிலிருந்து இந்திய மொழிபெயப்பு (தமிழ்-இந்தி) ஒழுங்குமுறையின் இரு கூறுகள் இவ்வாய்விற்கு எடுத்துக்கொள்ளப்பட்டுள்ளன என்பதன் அடிப்படையில் இவ்விரு செயல்பாடுகளையும் வெளிப்படுத்தும் பொருட்டுத் தமிழ் தரவுத் தொகுதியில் எவ்வாறு சொல்வகை அடையாளப்படுத்தலும் தொடர் பிரித்தலும் நடைபெறுகின்றன என்பதைக் கூறும் முகமாகக் கீழே எடுத்துக்காட்டு தரப்பட்டுள்ளது.

<NP> தலம்/NN :/SYM </NP> <NP> வைணவ/JJ தலம்/NN </NP>
<NP> அமைவிடம்/NN :/SYM </NP> <NP> திருச்சியிலிருந்து/NN </NP> <NP> 3/QC மைல்கள்/NN தூரம்/NN </NP>
<NP> திருக்கோயிலின்/NN பெயர்/NN :/SYM </NP> <NP> ஸ்ரீ/NNC அரங்கநாதசுவாமி/NNP திருக்கோயில்/NN </NP>
<NP> இறைவன்/NN /SYMஇறைவி/NN திருநாமங்கள்/NN :/SYM </NP> <NP> ஸ்ரீ/NNC அரங்கநாத/NNPC சுவாமி/NNP </NP> <NP> ஸ்ரீ/NNC ரங்கநாச்சியார்/NNP </NP>
<NP> தீர்த்தங்கள்/NN :/SYM </NP> <NP> சந்திரபுஷ்கரணி/NN </NP> <CCP> மற்றும்/CC </CCP> <NP> சூரிய/JJ புஷ்கரணி/NN </NP>
<NP> முக்கிய/JJ விழாக்கள்/NN :/SYM </NP> <NP> வைகுண்ட/NNC ஏகாதசி/NN பிரம்மோற்சவம்/NN </NP>
<NP> சிறப்பு/NN அம்சம்/NN :/SYM </NP> <NP> 108/QC வைணவத்/JJ திருத்தலங்களில்/NN </NP> <JJP> முதன்மையான/JJ </JJP>
<VGF>திருக்கோயிலாகும்/VM ./SYM </VGF>
<NP> 12/QC ஆழ்வார்களில்/NN </NP> <NP> 11/QC ஆழ்வார்கள்/NN </NP> <NP> இதன்/PRPபுகழை/NN </NP> <NP> பெருமையை/NN </NP> <NP> கீர்த்தியைப்/NN </NP> <VGF>பாடியுள்ளார்கள்/VM ./SYM </VGF>

<NP> இந்த/DEMதலத்தில்தான்/NN </NP> <NP> ஆண்டாள்/NNP </NP> <NP> திருப்பாணாழ்வார்/NNP </NP> <NP> பீபிநாச்சியார்/NNP </NP> <CCP> மற்றும்/CC </CCP> <NP> பலர்/NN </NP> <NP> அரங்கனின்/NNP திருவடியை/NN </NP> <VGF>அடைந்தனர்/VM ./SYM </VGF>

<RBP> சுயம்புவாக/RB </RBP> <NP> உருவான/JJ ஒரே/QC திருக்கோயில்/NN ./SYM </NP>

<NP> போக்குவரத்து/NN மையம்/NN :/SYM </NP> <NP> திருச்சி/NNP ./SYM </NP> <NP> அரங்கன்/NNP </NP> <VGNF> குடிகொண்டுள்ள/VM </VGNF> <NP> திருவரங்கம்/NNP </NP<NP> திருச்சி/NNP /SYM</NP> <NP> சென்னை/NNP கார்டு/NN ரெயில்/NN பாதையில்/NN </NP> <BLK> உள்ள/PSP </BLK> <NP> ரேVயில்/NN </NP> <VGF>நிலையமாகும்/VM ./SYM </VGF>

<NP> சென்னையிலிருந்து/NNP </NP> <NP> 207/QC மைல்கள்/NN தூரத்திலும்/NN </NP> <NP> திருச்சிராப்பள்ளி/NNP சந்திப்பிலிருந்து/NN </NP> <NP> 3/QC மைல்கள்/NN தொலைவிலும்/NN </NP> <VGF>உள்ளது/VM ./SYM </VGF>

<NP> திருவரங்கம்/NNP நகரம்/NN </NP> <NP> திருச்சிராப்பள்ளி/NNP மாநகரின்/NN புறநகரில்/NN </NP> <BLK> உள்ள/PSP </BLK> <JJP> சிறிய/JJ </JJP> <VGF>நகரமாகும்/VM ./SYM </VGF>

<NP> ஒருபுறம்/NN காவிரியும்/NNP </NP> <NP> மற்றொருபுறம்/NN கொள்ளிடமும்/NNP </NP> <VGNF>பாய்ந்தோடும்/VM </VGNF><NP> இந்த/DEMதிருத்தலத்துக்கு/NN </NP> <RBP> ஆண்டுமுழுவதும்/RB </RBP> <NP> ஆயிரக்கணக்கான/JJ இந்துக்கள்/NN </NP> <VGNF>வந்து/VM </VGNF><NP> அரங்கனை/NNP </NP> <VGF>வழிபட்டு/VM செல்கின்றனர்/VAUX ./SYM </VGF

<RBP> குறிப்பாக/RB </RBP> <NP> வைஷ்ணவர்களுக்கு/NN </NP> <NP> இதுPSP</NP> <JJP> முக்கியமான/JJ </JJP> <VGF>இடமாகும்/VM ./SYM </VGF>

7. முடிவுரை

இயந்திர மொழிபெயர்ப்பாளர்கள் ஒரு முழு தொடரியல் பகுப்பாய்வியை வலியுறுத்தாமல் ஒரு முழுமையற்ற தொடரியல் பகுப்பாய்வியை பரிந்துரைக்கின்றனர்.

தமிழ்சார்ந்த இயந்திர மொழிபெயர்ப்புக்கும் இத்தகைய முழுமையற்ற தொடரியல் பகுப்பாய்வி போதுமானதாகும். இத்தகைய பகுப்பாய்வி ஆழமில்லாப் பகுப்பாய்வி என அழைக்கப்படுகின்றது. இப்பகுப்பாய்வி செய்யும் செயல்பாடு தான் ஆழமில்லாப் பகுத்தலாகும் (shallow parsing). ஆழமில்லாப் பகுத்தலின் ஒரு பகுதியாக வரும் தொடர் பகுப்பாய்வு (Chunking) பற்றி இங்கு விரிவாகப் பேசபடுகின்றது.

துணைநூற்பட்டியல்

அகத்தியலிங்கம், ச. 2002. தமிழ் மொழி அமைப்பியல். சிதம்பரம்: மெய்யப்பன் பதிப்பகம்.

Aarts.J. and Meijs, W. (Eds.)1984. Corpus Linguistics: Recent Development in the use of Computer Corpora in English Language Research. Amsterdom-Atlanta. G.A.:Rodopi.

Aarts.J. and Meijs, W. (Eds.)1986. Corpus Linguistics II: New Studies in the Analysis and Explanation of Computer Corpora. Amsterdam-Atlanta, G.A.: Rodopi.

Aijmer, K. and Alwood, J. (Eds.) 2004. Dialogic Analysis: New Trends in Dialogue Analysis, Tubingen: Niemeyer.

Aijmer, K. and Altenberg, B. (Eds.) 1991. English Corpus Linguistics: Studies in Honour of Jab Stvartvik. London: Longman.

Akshar Bharati, Rajeev Sangal, Dipti Misra Sharma and Lakshmi Bai. 2006. AnnCorra: Annotating Corpora Guidelines for POS and Chunk Annotation for Indian Languages, Technical Report, Language Technologies Research Centre IIIT, Hyderabad.

Arulmozhi. P, Sobha L. 2006. A Hybrid POS Tagger for a Relatively Free Word Order Language. In proceedings of MSPIL-2006, Indian Institute of Technology, Bombay.

Atwell, E. (Eds.) 1993. Corpus Based Computational Linguistics. Amsterdam: Rodopi.

AU-KBC Post Tagset Tamil. AU-KBC Research Centre, MIT Campus of Anna University, Chrompet, Chennai-44.

Bernadini, S. 2000. Competence, Capacity, Corpora, Bologna: CLUEB.

Biber, D. Conrad, S. and Reppen, R. 1998. Corpus Linguistics-Investigating Language Structure and use. Cambridge University Press.

Chomsky, A.N. 1968. Language and Mind. New York: Harcourt Brace.

Chomsky, A.N. 1972. Studies in Generative Grammar. The Hague: Mouton.

Clear, Jereny, 1992. Corpus sampling in New Directions in English Language Corpora (ed.) Gerhars Leitner. Mouton de Gruyter, New york.

CIIL, HINDI, CORPORA Tagset. Finalized on August 29, 2006

Dash, N.S. 2005. Corpus Linguistics and Language Technology (with reference to Indian Languages). New Delhi: Mitial Publications.

Dhanalakshmi V, Anandkumar M, Vijaya M.S, Loganathan R, Soman K.P, Rajendran S, 2008. "Tamil Part-of-Speech tagger based on SVMTool", In Proceedings of the COLIPS International Conference on natural language processing (IALP), Chiang Mai, Thailand. 2008. P.No:

Dhanalakshmi V, Anand kumar M, Vijaya M.S, Loganathan R, Soman K.P, Rajendran S. 2008. "Tamil Part-of-Speech tagger based on SVMTool", Proceedings of International Conference on Asian Language Processing 2008 (IALP 2008), November 2008, Chiang Mai, Thailand.

Dhanalakshmi V., Padmavathy P., Anand Kumar M., Soman K.P., Rajendran S., 2009 "Chunker For Tamil Using Machine Learning" 7th International Conference on Natural Language Processing 2009 (ICON2009), IIIT Hyderabad, India.December 2009. P.No:

Dhanalakshmi.V, Anand kumar M, Rekha R U, Arun kumar C, Soman K P, Rajendran S. "Morphological Analyzer For Agglutinative Languages Using Machine Learning Approaches", Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, ARTCom 2009, Oct 2009 Kottayam, India .

Dhanalakshmi.V , Anand kumar M, Padmavathy P, Soman K P, Rajendran S, 2009 “Chunker For Tamil” Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, ARTCom 2009, October 2009, Kottayam, India.

Dhanalakshmi. V , Anand kumar M, Soman K P, Rajendran S , 2009 “Postagger And Chunker For Tamil Language” Proceedings of TAMIL INTERNET CONFERENCE 2009 , October 2009, Cologne, Germany.

Anand kumar M, Dhanalakshmi V, Soman K P, Rajendran S. 2009. “A Novel Approach For Tamil Morphological Analyzer”, Proceedings of TAMIL INTERNET CONFERENCE 2009, October 2009, Cologne, Germany.

Dhanalakshmi V., Anand Kumar M., Rekha R.U., Arun Kumar C., Soman K.P., Rajendran S., 2009."Morphological Analyzer for Agglutinative Languages Using Machine Learning Approaches," artcom, pp.433-435, 2009 International Conference on Advances in Recent Technologies in Communication and Computing, 2009,IEEE Press, doi: 10.1109/ARTCom.2009. P.No:184

Dhanalakshmi.V , Padmavathy P, Anand kumar M, Soman K P, Rajendran S, 2009 “Chunker For Tamil Using Machine Learning” 7th International Conference on Natural Language Processing 2009(ICON2009), December 2009, IIIT Hyderabad, India.

Dhanalakshmi V, Anand kumar M, Shivapratap G, Soman K.P, Rajendran S. 2009. “Tamil POS Tagging using Linear Programming” International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May-2009. ISSN 1797-9617.

Dhanalakshmi.V , Anand kumar M, Padmavathy P, Soman K P, Rajendran S. 2009. “Chunker For Tamil”, Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, ARTCom 2009, October 2009, Kottayam, India.

Eagles. Expert Advisory Group on Language Engineering Standards.

Eagles –Recommendations for the Morphosyntatic Annotaton of Corpora. Eagles Document EAG – TCWG-MACIR. Version of Mar, 1996.

Francis, W.N. and Kucera, H. 1964. Manual of Information to accompany A Standard Corpous of present day edited American English. Department of Linguistics, Brown Uniersity, USA.

Garside, R.Leech, G. and Mcenery, A/ (Eds.) 1997. Corpus Annotation Linguistics Information from computer text Corpora.London: Longman.

Gerbig, A.1997. Lexical and Grammatical Variation in a Corpus: A computer Assisted study of Discourse on the Environment. London: Peter Long Publishing.

Ghadassy, M., Henry.A., and Roseberry, R.L. (Eds) 2001. Small Corpus studies and ELT : Theory and practice Amsterdam / Philadelphia. John Benjamins.

Graf,D. 1996. Relative clauses in their Discourse Context : A Corpus – Based study Unpublished M.A., Thesis : Freiburg.

Granger, S. and Tyson, S.P. (Eds) 2003. Extending the scope of corpus- based Research: New Applications. New challenges. Amserdam: Rodopi.

Granger,S. Hung, J.And Tyson, S.P.Eds. 2002. Computer Learner Corpora, Second Language Acquistion and Foreign Language Teaching. Amsterdam: John Benjamins.

Greenbaun, S. (Ed.) 1996. Camparing English Worldwide: the International Corpus of English. Oxford: Clarendon.

Haan, Peter de, 1992. The optimum Corpus sample size? in New directions in English Language Corpora (ed.,) Mouton de Gruyter, New York.

Hunston, S. 2002, Corpora in Applied Linguistics, Cambridge University Press.

ILMT Consortium. 2008. IL MT system, Software Requirement Specification (SRS). July 2008, IIIT, Hyderabad.

Jayaram, B.D. 1996. Development of corpora in Indian languages: Problems and suggested solutions. Paper presented at workshop on Indian Language Corpus and its applications at CIIL, Mysore.

Jayaram, B.D Umarani P. 1998. Grammatical Category Disambiguation: A Probabilistic Model in South Asian Language Review, Vol. vII. No.1. Jayaram, B.D and Rajyashree, K.S corpora in Indian Language. Tamil Text

Johansson, S. and Hofland, K. (Eds) 1982. Computer corpora in English Language Research. Bergen: Nowegian computing Center for the Humanities.

Johansson, S. and Stenstrom, A.B. (Eds.) 1991. English computer Corpora: Selected Paper and Research Guide. Berlin: Mouton de Gruyter.

Kennedy, G. 1998. An Introduction to Corpus Linguistics. New York: Addison-Wesley Longman Inc.

Kettemann, C.B. and Marko, G. (Eds.) 2002. Teaching and Learning by Doing Corpus Analysis. Language and Computers: Studies in Practical Linguistics 42. Amsterdam-Atlanta G.A. Rodopi.

Kirk, J.M. (Ed.) 2000. Corpora Galore: Analysis and Techniques in Describing English, Amserdam, Atlanda, G.A. Rodopi.

Kyto, M. Thalainen, O.and Rissanen, M. (Eds.) 1998. Corpus Linguitics, hard and Soft : Proceedings of the 8 th International Conference on English Language Research on Computerised Corpora. Amsterdam: Rodopi.

Kumar, R. and S. Viswanathan, “tamil Morphological Analyser” at http:/www.av. KBC.org research-areas/projects/documentation.html (17.9.2001)

Lancushine, I.E Carol and C.F. Meyer (Eds.) 1997. Synchronic Corpus Linguistics. Bergen, Norway: ICAME.

Ljung, M. (Ed.) 1997. Corpus-Based Studies in English. Papers from the 17 th International Conference on English Language Research Based on Computerized Corpora. Amsterdam:Rodopi.

Lehmann. 1989. the Grammar of Modern Tamil. Pondicherry Institute of Linguistics and Culture. Pondicherry.

Mathhews, P.H. 1972. Inflectional Morphology. Cambridge University Press, London.

Mair, C. and hundt, M. (Eds.) 2000. Corpus Linguistics and Linguistics Theory. Amsterdam-Atlanda, GA: Rodopi.

Mecnery, T. and Wilson, A. 1996. Corpus Linguistics. Edinburg: Edinburgh University Press.

Mcenery, T. Rayson, P.and Wilson, A. (Eds.) 2002. A Rainbow of Corpora: Corpus Linguistics and the Language of the World. Munchen: Licolm Europe.

Meyer, C.F. 2002 English Corpus Linguistics. Cambridge University Press.

Nelson, G., Wallis, S. and Aarts, B. 2002 Exploring Natural Language: Working With the British component of the International Corpus of English. Amsterdam/Philadelphic John Benjamins.

Ookes, M.p. 1998. Statistics for Corpus linguistics. Edinburgh: Edinburg UniversityPress.

Ooi, V.B.V.1997. Computer Corpus Lexicogrphy. Edinburgh: Edinburg University Press.

Oostdijk, N.and Dehann, P. (Eds.) 1994. Corpus Based Research into Language Amsterdam-Atlanda, GA: Rodopi

Partington, A. 1998. Patterns and meanings-Using Corpora for English Language Reseach and Teaching. Amsterdam, Philadelphia: John BEnjamins.

Percy, C., Meyer, C.F., and Lancashire, I. (Eds.) 1996. Synchronic Corpus Linguistics, Amsterdam – Atlanta, G.A: Rodopi.

Peters, B.P., Collins, P. and Smith, A. (Eds.) 2002. New Frontiers of Corpus Research Language and Computers. Amsterdam – Atlanta, G.A. Rodopi.

Rajendran, S. 1995. "A Machine Bridge for Tamil." Proceedings of Eighth World Tamil Conference. Thanjavur.

Rajendran, S. 1997. "Grammatical Formalism and Computational Analysis of Nominal Compounds in Tamil." South Asian Language Review, vol 7, no.1., 27-46.

Rajendran.S. 1999. "Spell and Grammer Checker for Tamil." Paper read in 27th All Indian Conference of Dravidian Lingustics.

Rajendran.S. 2000. "Preliminaries to the Preparation of spell and Grammer checker for Tamil." Language in Indian WWW.languageindia.com.

Rajendran.S. 2006. "Parsing in Tamil: Present state of Art." Indian Linguistics: 67:159 – 67.

Rajendran, S., Arulmozi, S.Ramesh Kumar and S.Viswanathan. 2003. "Computational Morphology of Verbal Complex". In B.Ramakrishna Reddy (ed.) Word Structure in Dravidian. Kuppam: Dravidian University 376 – 398.

Rajendran, S. et al. 2003. "Computational Morphology of Verbal Complex." In: B. Ramakrishna Reddy (edited) Word Structure in Dravidian, Kuppam: Dravidian University, & Language in India 3:4, www.languageinindia.com, April 2003.

Ramaswamy, V. 2002. Morphological Generator for Tamil. Unpublished M.Phil Dissertation, University of Hyderabad, Hyderabad.

Ranganathan, 1997. “A Lexical phonology Approach Tamil word by Computer”. International Journal of Dravidian Linguistics, 26, 1:57 – 70.

Rekha, R U Anand kumar M, Dhanalakshmi.V , Soman K P, Rajendran S, 2010 “Morphological generator for Tamil a new data driven approach”, Tamil Internet Conference 2010, June 2010, Cemmozhi maanaadu, Coimbatore, India.

Sengupta, G. 1997. “Three models of Morphological processing” South Asian Language Review, 8:1, 1-26.

Shanmugam, C. 2001. “Computer Analysis of Simple sentence in Tamil”. Paper read in UGC SAP National Seminar on Computational Linguistics and Dravidian Languages, 22-24 February, 2001, CAS in Linguistics, Annamalai University, AnnamalaiNagar.

Shanmugam, C. 2002. “Grammer and Parser: A program for Syntactic Parsing in Tamil” International Seminar on Tamil Computing 27 -28 February and March 1, 2002. University of Madras: Chennai.

Sivashanmugam, C. 1998. “Language Technology and Morphological Processing”. The Administrator, 45:3.

Sivashanmugam, C. 2000. “A Model for computer Analysis of Verb in Tamil” In working papers in Linguistics, Department of Linguistics, Bharathiyar University, Coimbatore.

Sivashanmugam, C. 2002. “Morphological: processor for Negative Constructions in Tamil” Indian Conference on Natural Language processing, Anna University, Chennai.

Selkairk, E.O. 1983. The Syntax of words. Cambridge, Mass: MIT Press.

Selting, M.and Couper. Kuhlen E. (Eds.) 2001. Studies in International Linguistics. Amsterdam / Philadelphia: John Benjamins.

Sharman, R. 1990. Hidden Markov Model Methods for word Tagging. Report 214. Winchester : IBM UK Scientific centre.

Sinclair, J. 1991. Corpus, Concordance, Collocation. Oxford University Press

Souter, C. and Atwell, E. (Eds.) 1993, Corpus based computational Linguistics. Amsterdam: Rodopi.

Stenstorm, A.B. Anderson, G.and Hasund, I.K.2002. Trends n Teenage Talk: Corpus Compilation, Analysis and Findings, Amsterdam: John Benjamins.

Stubbs, M. 1996. Text and corpus Analysis : Computer Assisted Studies of Lanaguage and culture. Oxford : Balckwell Publishers.

Summers, D. 1991. Longman/Lancaster English Language Corpus: criteria and Design Harlow: Longman.

Svartvik, J, (Ed.) 1990. The London Corpus of spoken English: Description and Research. Lund Studies in English 82, Lund: Lund University Press.

Starvik, J. Ed. 1992. Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 – Stockholm, 4-8 August 1991. Berling, New York, Mouton De Gruyer.

Thomas, J. and Short, M. (Eds.) 1996. Using Corpora for Language Research: Studies in the Honour of Geoffery Leech. London and New York: Addison Welsely Longman.

Tognini–Boneli. E.2001. Corpus Linguistics at work. Amsterdam: John Benjammins.

Umarani, P. 2001. Morphological Analyser for Tamil. Unpublished Doctoral Dissertation.

Vaishnavi Ramaswamy. 2000. A Morphological Generator for Tamil. M.Phil. Thesis, CALTS. University of Hyderabad, Hyderabad.

Veronis, T. (Ed.) 2000. Parallel Text Processing. Alignment and use of Translation Corpors. Dordreeht: Kluwer Academic Publishers.

Wichmann, A. Fligelstone, S. Mcenery, T. and Knowles, G (Eds.) 1997. Teaching and Language Corpora, London: Longman

Young, S. and G.Bloothooft (Eds.) 1997. Corpus based methods in Language and Speech Processing, Vol. II, Dordrecht: Kluwer Academic Press.